A Statistical Evaluation of Hybrid Potato Breeding

James R. Adams



Propositions

1. The same statistical models used in other hybrid crops are amenable to hybrid potato
   (this thesis)

2. Starting a new hybrid breeding program without quantitative genetic theory is unlikely to succeed
   (this thesis)

3. Potato is just a vegetable masquerading as a field crop

4. It is professionally disqualifying to foster the adoption of generative AI in academia

5. Publicly funded research is a better model than public-private consortiums

6. Higher education institutions are responsible for upholding international law

7. Dutch sandwich culture is excessively frugal and passionless

Propositions belonging to the thesis, entitled
A Statistical Evaluation of Hybrid Potato Breeding
James Adams
Wageningen, October 22nd, 2026

A Statistical Evaluation of Hybrid Potato Breeding

James R. Adams

Thesis committee

Promotor:
Prof. Dr F.A. van Eeuwijk
Professor of Mathematical and Statistical Methods
Wageningen University & Research

Co-promotors:
Dr Michiel de Vries
Global Director Potato Science
McCain Foods, Canada

Dr Chaozhi Zheng
Senior Researcher Biometris
Wageningen University & Research

Other members:
Prof. Dr Mario Calus, Wageningen University
Dr Chris Maliepaard, Wageningen University
Dr Marlee Labroo, Bayer, United States of America
Dr Valentin Wimmer, Aardevo, Nagele

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

A Statistical Evaluation of Hybrid Potato Breeding

James R. Adams

Thesis
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof. Dr Carolien Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 22 October, 2026
at 3:30 p.m. in the Omnia building.

James Adams
A Statistical Evaluation of Hybrid Potato Breeding
172 pages.

PhD thesis, Wageningen University, Wageningen, The Netherlands (2026)
With references, with summary in English

ISBN: 978-94-6534-731-8
DOI: 10.18174/681079

”감자엿이 제일이래
많이 들고 장수하소 장수령감
감자술잔 비우면서 하는 말이”

”The potato taffy is the best they say
the old man who ate a lot and lived to a ripe old age
says while emptying a glass of potato liquor . . .”
- Excerpt from ”Potato Pride” (감자자랑)

Contents

Page

Contents vii

Chapter 1 Introduction 1

Chapter 2 Little heterosis found in Diploid Hybrid Potato — The genetic underpinnings of a new hybrid crop 21

Chapter 3 Efficient genomic prediction of yield and dry matter in hybrid potato 53

Chapter 4 Digging up QTL: Evaluating the efficiency of marker-assisted selection in hybrid potato 79

Chapter 5 Evaluating Multiallelism for Genomic Prediction Applications in Potato 101

Chapter 6 General Discussion 131

Summary 153

About the author 157

Acknowledgements 160

1

Introduction

1.1 Potato: a very brief history

Potato (*Solanum tuberosum*) has transformed the diet of billions of people over the last century. Cultivated for its starchy tubers, this vegetable has become the most important non-cereal food crop making it a keystone in global food security[1]. Originating from modern-day Peru, potato has a complex evolutionary history which is still under intense scrutiny. Tuber bearing *Solanum* species (known as *Petota*), represent one of the most diverse sections within the genus, comprising over 200 species with varying levels of ploidy and ecological adaptations[2]. Only recently has the potato lineage been found to have originated as an inter-specific hybrid between Etuberosum and tomato lineages upwards of 9 million years ago[3]. Neither of these ancestral species were tuber-bearing, and only through subsequent evolutionary events did tuber formation emerge as a key adaptation for reproduction in high-altitude environments[4]. This hybridization event precipitated an unprecedented radiation resulting in over 100 tuber-bearing species distributed across the Andes stretching from modern-day Colombia to Bolivia[2]. This diversification was possible through the incredible adaptation of these species to multiple climates and ecological niches, ranging from high-altitude páramo ecosystems to subtropical valleys.

Domestication of potato is thought to have originated around 10,000 years ago in the Lake Titicaca region of the Andes, but like its early history, aspects of potato's domestication are rather muddled[5]. There are two primary origin theories, namely, a single domestication event from *Solanum bukasovii* or multiple domestication events from different wild progenitors. The impact of domestication on potato did not only result in differences in the harvested tubers (e.g. enlargement of tubers, reduction of anti-nutrients content like glycoalkaloids, and changes in tuber shape and skin color), but also altered flowering patterns, photoperiod sensitivity, and reproductive biology[4]. Through domestication, potato evolved from a short-day adapted plant to one with greater environmental plasticity. The introduction of tetraploid Andean landraces to Europe in the mid-16th century marked the beginning of potato cultivation in European agriculture. It was not until the introduction and admixture with Chilean potato landraces that potato was adapted to long-day conditions, leading to the widespread adoption and spread of potato across Europe[6,7].

The origins of selective efforts of potato in Europe, like many crops, were quite rudimentary. Just as potato was being adopted by the peasantry in England and Ireland in the mid-18th century, a few prescient observations were made around their propagation. The growing of potato seedlings over seed tubers was thought to buffer the deterioration observed in extant clones, perhaps observing the accumulation of pathogen load over multiple clonal generations[8]. Additionally, while a few different "*types*" of potato were observed to be grown in Britain, the reliance on clonal propagation meant that European potato agriculture was one of the first true genetic monocultures with devastating consequences. Between 1845 and 1849, Europe experienced one of the worst famines in modern history. Due to a continent-wide *Phytophthora infestans* epidemic, further compounded by an exploitive farmer tenant system that encouraged nearly exclusive potato cultivation, catastrophic crop failure fell upon Ireland. The resulting Great Hunger in Ireland

and related Continental Famine in Europe was responsible for an excess deaths of over 1.1 million people[9]. While the exact causes of the blight were not rightly known at the time, it was generally understood that there was a *deficiency* in this new crop and that any human dependence on it necessitated immediate improvement.

1.2 Modern Potato breeding

By the 19th century, potato cultivation was expanding across North America. Moreover, intentional crosses and selection of well-adapted seedlings was becoming commonplace leading to the creation of multiple varieties in the latter decades preceding the 20th century[10]. This period of breeding was prolific and marked by the influx of exotic germplasm from Chile such as Rough purple Chili (˜1850) or Daber (˜1860)[8,11]. The progeny of these varieties have been so valuable that they continue to be cultivated even now; these include varieties such as Russet Burbank (1908), King Edward (1902), and Bintje (1910)[12]. These new varieties were not only carrying valuable resistances to blight and scab, but were also being bred for their valuable starch to satisfy the burgeoning North American and European fermentation and processing industries[13]. Potato's newfound industrial capacity was taking place just as its popularity in its fried form, *pommes frites*, was being discovered on a new world stage. Already by the early 1920s, potato production was feeding a multi-faceted global economy as it was being formed in real time.

Once the province of hobbyists, selection of novel and better potato varieties began rapid formalization as breeders began to select for well-adapted clones through the early 20th century. By 1950, public breeding institutions were founded in the United States (USDA breeding institutes; 1910), Britain (Scotland's plant breeding station; 1920), India (Potato Breeding Station; 1935), and China (1940)[14–17]. For many of these institutions, the primary end was adapting potato to important production locations, but also for the improvement of traits like maturity, marketable yield fractions, insect and disease resistance, external tuber quality (e.g. eye depth, skin cracking), and cooking quality to name a few[18,19]. In parallel to this, international gene banks were established and charged not only with the task of taxonomical classification, but also assessing the genetic utility of these collections. The International Potato Center (a member of the Consultative Group on International Agricultural Research centers) was one of the largest of these institutions but other notable germplasm centers includes the United States Potato Introduction Station, BrACySol (INRAE), and the Vavilov Institute of Plant Genetic Resources[20–22]. These centers have become invaluable to modern potato improvement through their identification of novel abiotic and biotic resistances among exotic germplasm.

Over the past 50 years, potato research benefited from rapid methodological advances in both molecular and cell biology. Potato was one of the first crop plants to utilize *in vitro* propagation and manipulation techniques. This includes the development of protocols related to minituber propagation & virus elimination[23,24]. Along with having a positive impact on production of clean potato stock, *in vitro* methods made potato a prime candidate for gene-editing technologies. By the turn of the century, potato was on the forefront of biotechnological improvement with several transgenic clones targeting traits such as in-

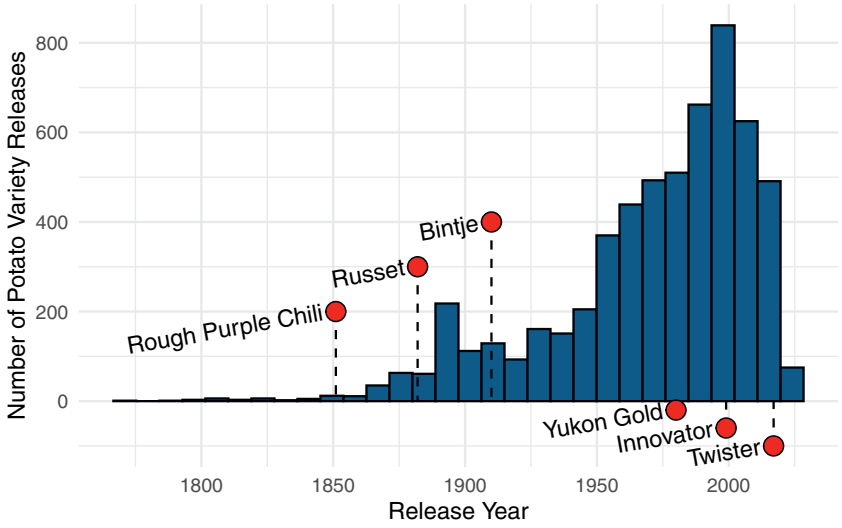


Figure 1.1: Number of released cultivars globally since 1750 where the year of release and pedigree is known (van Berloo et al., 2007).

sect resistance and amylose fractions[25,26]. While gene-editing technologies are currently under-utilized among breeders in western Europe, there are many other agronomic and tuber composition traits that would be worthwhile candidates for transformation[27]. Perhaps more than any other recent technology, the development and application of molecular markers for simple and complex trait associations have greatly impacted the efficiency of potato breeding[28]. Some of the first genetic mapping efforts with restriction and amplified fragment length polymorphisms (RFLP and AFLP, respectively) were responsible for the initial mapping for regions providing resistance to *Phytophthora infestans*, Potato Virus X, and *Globodera rostochiensis*[29–31]. Almost exclusively, these first genetic maps were developed in diploid mapping populations, but with the advent of the next generation sequencing and the creation of the first reference genome, linkage mapping efforts soon expanded to tetraploid populations[32–34]. This has led to incredible progress in both quantitative trait locus (QTL) mapping methodology as well as genomic prediction opening up many modern data-driven tooling to potato[35–38].

Challenges in potato genetic improvement

Much of potato's history chronicles difficulties surrounding potato's cultivation. Even incidental aspects of potato's evolution reach through millennia and affect growers with many ramifications on modern production systems. Characteristics of potato's biology also hamper selection and improvement of potato germplasm; these I highlight below.

While species within *Petota* range widely in ploidy number, cultivated potato is almost entirely tetraploid ($2n = 4x = 48$) exhibiting a tetrasomic pattern of inheritance. Potato's autoploidy naturally has some ramifications for breeding. Polyploidy in crops is often described as a sort of double-edged sword in the context of genetic improvement. Multiple

genome copies is generally associated with several benefits in crop species including larger organ size, greater *heterotic* vigour, and more robust performance to sub-adequate climate conditions[39,40]. These benefits often arise as a response to genome duplication followed by gene-conversion or gene-neofunctionalization which is thought to be a beneficial mode for positive evolutionary selection. Some researchers would describe *Petota*'s polyploidy as one of the major vehicles for its wide environmental adaptability[41,42].

Despite these benefits in an evolutionary setting, polyploidy through the lens of breeding is generally seen as cumbersome. Gene frequencies in tetraploids are strongly buffered from perturbations, making directional selection slower in a tetraploid system relative to their diploid counterparts[39]. This becomes quite clear when considering a single locus model. Using the single locus model[43], we can define the allelic variance at a given locus as the variance of a binomial distribution, i.e.

$$\sigma_p^2 = \frac{p \cdot (1-p)}{l \cdot N}$$

where p is the allele frequency in the population, l is the ploidy number and N is the number of individuals in the population. Comparing two populations (one diploid, $l = 2$, and the other tetraploid, $l = 4$) comprised of 10 individuals ($N = 10$) and an allele frequency of 0.35 ($p = 0.35$), the allelic variance would be 0.011 and 0.0057 in the diploid and tetraploid populations, respectively. Considering how these allelic variances affect the probability of allelic shifts, we can examine the question of a random positive 0.1 shift in the next generation. Evaluating this probability ($P(p \geq 0.45) = 1 - \Phi\left(\frac{0.1}{\sigma_p}\right)$) in both populations results in a probabilities of 0.17 and 0.09 in the diploid and tetraploid populations, respectively. This demonstrates that gene frequencies even under a simple drift model are buffered from random fluctuations in polyploids relative to their diploid counterparts. This touches on all population dynamics which artificial selection seeks to modulate including inbreeding, single locus selection and fixation, and even population improvement schemas which extend across thousands of loci.

Speed of selection aside, potato also carries with it the burden of deleterious mutation load. Like many other polyploid crops, accumulation of deleterious mutations was a byproduct of a genome duplication event leading to inherent redundancy of multiple genome copies. This genome masking and clonal propagation strategy resulted in the amassing of mutations as maternal plants passed on their genome to daughter tubers over thousands of generations[44]. This process was nearly universal in potato resulting in the hundreds of thousands of functionally deleterious mutations among both landraces and cultivated species. The impact of this evolutionary process is directly observable via inbreeding of tetraploid populations or haploidization of tetraploids leading to immediate loss of vigour, severe segregation distortion, and detectable zygotic and gametic selection signatures[45,46]. The moment a breeder sexually propagates tetraploid potato and attempts artificial selection among the progeny, the millennia of mutations are then unmasked resulting in incredible phenotypical variances in this first generation.

These large variances observed in segregating tetraploid populations have knock-on ef-

fects for artificial selection on tetraploid potato. First, a breeder must generate enough progeny per biparental cross to sufficiently "*sample*" from all possible genetic combinations possible to increase the likelihood of creating a progeny of higher merit for some trait(s). However, this presents a second challenge of *finding* that progeny of higher merit among hundreds of thousands of seedlings. Like any other field crop, the best means of evaluating a potato clone's merit is through multiple variety trials which would allow for an unbiased evaluation of those traits[47]. This on one hand sounds like a simple problem, but this gets into several issues unique to clonal crops. In the initial generations, breeders are reliant on true potato seed (TPS) based evaluation which excludes any possibility of replicated evaluations as these seedlings are being generated from segregating families. More problematic, however, is that phenotypic correlations between initial seedling trials and subsequent clonal generations are considerably low[48,49]. This limits most phenotypic based selection on tuber and yield characteristics until later clonal generations where sufficient tuber-based propagule is available. The cumulative effect of these factors is a reduced selection accuracy with slower genetic progress relative to other crops[50].

One of the more underappreciated challenges in potato breeding is the sheer number of traits which beset breeders. Depending on the program and market segment, a potato breeder could be considering over 40 traits among a panel of candidates[51]. Many of these characteristics also exhibit complex genotype by environment and even genotype by management interactions further necessitating high quality and consistent multi-environment trials. Specific economic traits such as starch content, cold sweetening, and skin firmness are especially sensitive to many post-harvest conditions[52]. This means that trialing of potato extends outside the field trial and includes the assessment of management factors across the haulming, harvest, and tuber storage stages.

Perhaps the most important category of traits in potato are those related to yield. These are expressed in terms of sheer mass (e.g. gross tonnes per hectare, marketable yield per hectare, tonnes of starch) or with respect to some component of yield (e.g. number of tubers per stem, average size of tuber) and are often the direct targets of evaluation and selection[53–55]. One of the most pressing questions for potato breeders is with regard to increasing yield in potato and why there appears to be less genetic improvement with respect to yield in other crops over the past century. Part of the complexity in this question is related to the aforementioned large trait surface in potato, but also potato's market segmentation. Because potato is divided into multiple distinct industries (ware, chips, processing, ingredient), there are multiple selection criteria for each which cannot be met with one program. So even if marketable yield had increased for russet types or tuber appearance for the chipping market, these gains are not easily transferable to other segments[19,56]. So while the rate of varietal release has accelerated over the past 40 years (Figure 1.1), this does not reflect an increase in yield fractions across market segments. What this demonstrates is a stratification in the improvement of potato which acts as an additional barrier for complex trait improvement.

1.3 Hybrid Potato breeding

Despite many modern advances in potato breeding[57], the traditional clonal breeding strategy has essential barriers that retard the breeding process. Many modifications of this schema have been proposed to circumvent these challenges, but one of particular interest is adapting potato to a hybrid breeding schema[17,58,59]. Potato is certainly not the first crop where this has been considered; maize aside, sugar beet, wheat, barley, and rice are all field crops which have viable F1 hybrid programs[60]. The justifications for such an endeavour are well known, but the benefits of hybridization in crops are comprised of factors economical, biological, and logistical. Focusing on the latter two components, hybrid breeding is designed around maximizing the heterotic response among the F1 progeny of inbred lines from opposing pools. While more capital intensive with longer breeding cycles than line breeding, hybrid breeding is characterized by its increased uniformity, exploitation of hybrid vigour, and capacity to quickly develop crop varieties[61,62]. Moreover, a hybrid breeding schema offers particular efficiencies for potato. First, is the direct utilization of true potato seed (TPS) in the trialling and release of cultivars[63,64]. This leverages TPS as a means of rapid propagule multiplication which further enables replicated field testing earlier in the breeding cycle. This directly affects the quality of phenotypic selection by increasing the heritability and thus allows for selection earlier in the breeding program. Along with this, hybrid breeding designs can be quite conducive for simultaneous improvement across multiple traits, especially through the incorporation of technologies like genomic selection[65,66]. Hybrid breeding programs are compartmentalized by design and allow for selection targets per pool thus decreasing the complexity of the large trait surface.

Benefits aside, transforming clonal potato into a hybrid crop is a resource intensive process that involves several prerequisite steps. Such an adaptation necessitates a few major measures to render a potential breeding program: (1) Breeding potato at a diploid level and (2) generation of self-compatible and self-fertile germplasm. Regarding ploidy, much has already been written here. Since the mid-20$^{\text{th}}$ century, there have been multiple efforts to breed with diploid germplasm. Interest in diploid breeding range from faster inbreeding rates, increased ease of introgressing exotic diploid germplasm, and bridge programs[18,67,68]. It should be said that while hybrid breeding is technically possible in autotetraploid potato, it is impractical for many of the aforementioned reasons. The major bottleneck with diploid breeding lies in the process of dihaploidization which is most commonly genome reduction via angiogenesis[69]. The other imperative in hybrid breeding is the ability to rapidly develop pure lines via selfing. Potato, practically an obligate outcrosser, presents a few barriers with its self-incompatibility mechanism being the most significant. Self-incompatibility in potato, while not universal, is widespread in tuber-bearing Solanum predominately in the form of gametophytic self-incompatibility locus, *Sli*[70,71]. Since its discovery, however, this locus is now one of the primary tools used in research and breeding programs to create self-compatible diploid germplasm[72,73]. While this is the primary barrier for inbreeding potato, symptoms of inbreeding depression in potato include fertility anomalies such as bud abortion, lower pollen count, and poor seed yields all of which have ramifications in hybrid breeding programs as well[74]. De-

spite these challenges, there are (at the time of writing) several active public and private programs with nearly or fully inbred diploid potato populations which are actively under development[46,75,76].

With many of these biological barriers addressed, we are still left with many questions surrounding the feasibility of hybrid potato programs. There is particularly little known regarding the population or genetic properties of current extant material. This is pertinent as there are few factors more relevant to the viability of a hybrid breeding program than evaluating the current status of complex traits. This includes the following: a thorough examination of heterotic vigour, current yield potential, and an assessment of relevant genetic models applicable to inference and prediction of field performance of hybrid potato. As we have already addressed, hybrid breeding should theoretically provide a means of monitoring and exploiting heterotic vigour as a breeding program progresses over time; if this vigour is less than expectation, then one of the primary benefits of such a system is unfulfilled for potato. Related to this, utilizing biometrical methods, we should be able to estimate accurate genetic parameters for yield components which offer both a vetting of the current capacity of F1 hybrids but also a forecast of the future potential of such programs[77]. Lastly, the creation of a new hybrid crop offers an opportunity to appraise the utility of novel biometrical methodologies for the prediction of genetic merit in a young hybrid breeding program.

1.4 Aim of this thesis

There remains a significant gap in comprehensive biometrical analyses of its genetic architecture and the application of genomic methods in hybrid potato. With the exception of a handful of published studies[17,58,78], there have been few quantitative inspections of the genetic architecture or surveys of genomic methods applied to hybrid potato. This thesis aims to address this gap through two primary directives. First, I wish to apply a battery of traditional genetic models successfully leveraged in other hybrid crops to evaluate the performance of a F1 hybrid potato and assess genetic mechanisms underlying multiple yield-related phenotypes. Second, I seek to use this material to evaluate modern predictive modeling frameworks to determine which technologies are most effective for hybrid potato breeding programs. The layout of this of thesis is as follows:

Chapter 2 begins with the evaluation of an F1 hybrid multi-environmental trial (MET). This chapter focuses primarily on estimating intra-trial spatial variation and statistically delineating several genetic components of variation, namely additive and non-additive genetic effects operating within this hybrid crossing block. I consider how these genetic fractions for several tuber traits might impact breeding strategies and selection decisions.

Chapter 3 examines the potential of multiple METs for training genome-wide regression models. This analysis is conducted on several heritable traits, with specific attention to different genetic parameterizations to assess the prediction accuracy of each model. The genomic models tested are analogous to those examined in **Chapter 2**, with the addition of a model validation procedure specifically designed for F1 hybrid datasets.

Chapter 4 provides a broader evaluation of genomic models and assesses their potential in several selection scenarios. Specifically, the efficiency of genomic selection is contrasted with marker-assisted selection (MAS) schemas based on putative quantitative trait loci (QTL) in a panel of F1 hybrids. The primary aim is to evaluate whether quantitative trait improvement in diploid hybrid potato could be more efficient with MAS-based procedures compared to genomic selection approaches.

Chapter 5 addresses the question of molecular marker information and whether multiallelic marker data offers any utility in predictive applications. Two forms of multiallelic information are considered: short-range haplotypes (or haplotags) and founder-origin variants (or identity-by-descent based variants). I examine the utility of both forms of multiallelism through a broad survey of predictive genomic models to assess their benefits over traditional SNP-based predictors.

Chapter 6 concludes with a synthesis of the previous experimental chapters, with particular emphasis on how these findings advance biometrical applications in hybrid potato. We conclude by offering a survey of currently under-served research topics and discussing their relevance to hybrid potato breeding systems, providing direction for future investigations in this emerging field.

Bibliography

1. Zaheer K and Akhtar MH. Potato Production, Usage, and Nutrition—A Review. Critical Reviews in Food Science and Nutrition. 2016 Apr 3; 56:711–21. DOI: 10.1080/10408398.2012.724479. PMID: 24925679. Available from: https://doi.org/10.1080/10408398.2012.724479 [Accessed on: 2025 Oct 19]
2. Spooner DM, Ghislain M, Simon R, Jansky SH, and Gavrilenko T. Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. Botanical Review. 2014; 80:283–383. DOI: 10.1007/s12229-014-9146-y
3. Zhang Z, Zhang P, Ding Y, Wang Z, Ma Z, Gagnon E, Jia Y, Cheng L, Bao Z, Liu Z, Wu Y, Hu Y, Lian Q, Lin W, Wang N, Ye K, Wang H, Zhang J, Zhou Y, Liu L, Li S, Lucas WJ, Särkinen T, Knapp S, Rieseberg LH, Liu J, and Huang S. Ancient Hybridization Underlies Tuberization and Radiation of the Potato Lineage. Cell. 2025 Sep 18; 188:5249–5265.e15. DOI: 10.1016/j.cell.2025.06.034. PMID: 40749684. Available from: https://www.cell.com/cell/abstract/S0092-8674(25)00736-6 [Accessed on: 2025 Oct 11]
4. Hardigan MA, Parker F, Laimbeer E, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS, Farré EM, Veilleux RE, and Robin Buell C. Genome Diversity of Tuber-Bearing Solanum Uncovers Complex Evolutionary History and Targets of Domestication in the Cultivated Potato. Proceedings of the National Academy of Sciences of the United States of America. 2017; 114. DOI: 10.1073/pnas.1714380114. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1714380114 [Accessed on: 2019 Feb 14]
5. Spooner DM, McLean K, Ramsay G, Waugh R, and Bryan GJ. A Single Domestication for Potato Based on Multilocus Amplified Fragment Length Polymorphism Genotyping. Proceedings of the National Academy of Sciences of the United States of America. 2005 Oct 11; 102:14694–9. DOI: 10.1073/pnas.0507400102
6. De Jong H. Impact of the Potato on Society. American Journal of Potato Research. 2016 Oct 1; 93:415–29. DOI: 10.1007/s12230-016-9529-1. Available from: https://doi.org/10.1007/s12230-016-9529-1 [Accessed on: 2025 Oct 11]
7. Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Luis Fernández-Alonso J, Prat S, and Burbano HA. The Origins and Adaptation of European Potatoes Reconstructed from Historical Genomes. Nature Ecology & Evolution. 2019 Jul; 3:1093–101. DOI: 10.1038/s41559-019-0921-3. Available from: https://www.nature.com/articles/s41559-019-0921-3 [Accessed on: 2025 Oct 11]
8. Glendinning DR. Potato Introductions and Breeding up to the Early 20th Century. New Phytologist. 1983; 94:479–505. DOI: 10.1111/j.1469-8137.1983.tb03460.x.

Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1983.tb03460.x [Accessed on: 2025 Dec 6]

9. Zadoks JC. The Potato Murrain on the European Continent and the Revolutions of 1848. Potato Research. 2008 Mar 1; 51:5–45. DOI: 10.1007/s11540-008-9091-4. Available from: https://doi.org/10.1007/s11540-008-9091-4 [Accessed on: 2025 Dec 21]
10. Stevenson F and Clark C. Breeding and Genetics in Potato Improvement. *Yearbook of Agriculture*. U.S. Government Printing Office, 1937 :405–44. Google Books: gs9gQTvtpFMC
11. Plaisted RL and Hoopes RW. The Past Record and Future Prospects for the Use of Exotic Potato Germplasm. American Potato Journal. 1989 Oct 1; 66:603–27. DOI: 10.1007/BF02853982. Available from: https://doi.org/10.1007/BF02853982 [Accessed on: 2026 Jan 10]
12. Van Berloo R, Hutten RCB, van Eck HJ, and Visser RGF. An Online Potato Pedigree Database Resource. Potato Research. 2007 Jan 1; 50:45–57. DOI: 10.1007/s11540-007-9028-3. Available from: https://doi.org/10.1007/s11540-007-9028-3 [Accessed on: 2025 Dec 6]
13. Willard M. Potato Processing: Past, Present and Future. American Potato Journal. 1993 May 1; 70:405–18. DOI: 10.1007/BF02849120. Available from: https://doi.org/10.1007/BF02849120 [Accessed on: 2026 Jan 10]
14. Luthra SK, Pandey SK, Singh BP, Kang GS, Singh SV, and Pande PC. Potato Breeding in India. CPRI, Shimla Technical Bulletin. 2006 :90. Available from: https://www.researchgate.net/profile/S-Luthra-2/publication/339998947_Potato_Breeding_in_India-_Editors_SK_Luthra_VK_Gupta_JK_Tiwari_Vinod_Kumar_Vinay_Bhardwaj_Salej_Sood_Dalamu_RP_Kaur_Raj_Kumar_G_Vanishree_Devendra_Kumar_Priyank_Mhatre_and_SK_Chakrabarti_2020_Potato_B/links/5ef0b02b92851ce9e7faff8f/Potato-Breeding-in-India-Editors-SK-Luthra-VK-Gupta-JK-Tiwari-Vinod-Kumar-Vinay-Bhardwaj-Salej-Sood-Dalamu-RP-Kaur-Raj-Kumar-G-Vanishree-Devendra-Kumar-Priyank-Mhatre-and-SK-Chakrabarti-20.pdf [Accessed on: 2026 Jan 10]
15. Jansky SH, Jin LP, Xie KY, Xie CH, and Spooner DM. Potato Production and Breeding in China. Potato Research. 2009 Feb 1; 52:57–65. DOI: 10.1007/s11540-008-9121-2. Available from: https://doi.org/10.1007/s11540-008-9121-2 [Accessed on: 2026 Jan 10]
16. Bradshaw JE. Potato Breeding at the Scottish Plant Breeding Station and the Scottish Crop Research Institute: 1920–2008. Potato Research. 2009 May 1; 52:141–72. DOI: 10.1007/s11540-009-9126-5. Available from: https://doi.org/10.1007/s11540-009-9126-5 [Accessed on: 2026 Jan 10]
17. Jansky SH and Spooner DM. The Evolution of Potato Breeding. *Plant Breeding Reviews*. Ed. by Goldman I. 1st ed. Wiley, 2018 Feb 14:169–214. DOI: 10.1002/9781119414735.ch4. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9781119414735.ch4 [Accessed on: 2025 Dec 6]

18. Krantz FA. Potato Breeding Methods 3: A Suggested Procedure for Potato Breeding. 1946 Feb. HDL: 11299/204107. Available from: https://hdl.handle.net/11299/204107 [Accessed on: 2024 Dec 29]
19. Douches DS, Maas D, Science KJC, and 1996 U. Assessment of Potato Breeding Progress in the USA over the Last Century. Crop Science. 1996; 36:1544–52. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/36/6/CS0360061544 [Accessed on: 2019 Apr 29]
20. Huaman Z, Salas A, Gomez R, Panta A, and Toledo' J. Conservation of Potato Genetic Resources at CIP. CIP Circular. 1986; 14
21. Bamberg JB and Spooner DM. The United States Potato Introduction Station Herbarium. TAXON. 1994; 43:489–96. DOI: 10.1002/j.1996-8175.1994.tb04854.x. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1996-8175.1994.tb04854.x [Accessed on: 2026 Mar 13]
22. Dalamu, Sood S, Mangal V, Kumar V, Sharma A, and Bhardwaj V. Potato: Collection, Conservation, and Evaluation of Genetic Resources. *Vegetable Crops*. Springer, Singapore, 2025 :143–78. DOI: 10.1007/978-981-97-8949-8_4. Available from: https://link.springer.com/rwe/10.1007/978-981-97-8949-8_4 [Accessed on: 2026 Mar 13]
23. HUSSEY G and STACEY NJ. In Vitro Propagation of Potato (Solanum Tuberosum L.) Annals of Botany. 1981 Dec 1; 48:787–96. DOI: 10.1093/oxfordjournals.aob.a086186. Available from: https://doi.org/10.1093/oxfordjournals.aob.a086186 [Accessed on: 2026 Mar 13]
24. Jones MGK. In Vitro Culture of Potato. *Plant Cell and Tissue Culture*. Ed. by Vasil IK and Thorpe TA. Dordrecht: Springer Netherlands, 1994 :363–78. DOI: 10.1007/978-94-017-2681-8_15. Available from: https://doi.org/10.1007/978-94-017-2681-8_15 [Accessed on: 2026 Mar 13]
25. Thornton M. The Rise and Fall of NewLeaf Potatoes. NABC Rep. 2003; 15:235–43. Available from: https://ecommons.cornell.edu/bitstreams/67589c53-9a17-466d-aa46-d3d6894ccea5/download [Accessed on: 2026 Mar 13]
26. Karlsson ME, Leeman AM, Björck IME, and Eliasson AC. Some Physical and Nutritional Characteristics of Genetically Modified Potatoes Varying in Amylose/Amylopectin Ratios. Food Chemistry. 2007 Jan 1; 100:136–46. DOI: 10.1016/j.foodchem.2005.09.032. Available from: https://www.sciencedirect.com/science/article/pii/S0308814605008101 [Accessed on: 2026 Mar 13]
27. Millam S. Developments in Transgenic Biology and the Genetic Engineering of Useful Traits. *Potato Biology and Biotechnology*. Elsevier Science B.V., 2007 Jan 1:669–86. DOI: 10.1016/B978-044451018-1/50072-5. Available from: https://www.sciencedirect.com/science/chapter/edited-volume/pii/B9780444510181500725 [Accessed on: 2026 Mar 13]
28. Gebhardt C. Molecular Markers, Maps and Population Genetics. *Potato Biology and Biotechnology*. Elsevier Science B.V., 2007 Jan 1:77–89. DOI: 10.1016/B978-044451018-1/50047-6. Available from: https://www.sciencedirect.com/science/chapter/edited-volume/pii/B9780444510181500476 [Accessed on: 2026 Mar 13]

29. Gebhardt C, Ritter E, Debener T, Schachtschabel U, Walkemeier B, Uhrig H, and Salamini F. RFLP Analysis and Linkage Mapping in Solanum Tuberosum. Theoretical and Applied Genetics. 1989 Jul 1; 78:65–75. DOI: 10.1007/BF00299755. Available from: https://doi.org/10.1007/BF00299755 [Accessed on: 2026 Mar 13]
30. Gebhardt C, Mugniery D, Ritter E, Salamini E, and Bonnel E. Identification of RFLP Markers Closely Linked to the HI Gene Conferring Resistance to Globodera Rostochiensis in Potato. Theoretical and Applied Genetics. 1993; 85:541–4
31. Meksem K, Leister D, Peleman J, Zabeau M, Salamini F, and Gebhardt C. A High-Resolution Map of the Vicinity of the R1 Locus on Chromosome V of Potato Based on RFLP and AFLP Markers. Molecular and General Genetics MGG. 1995 Jan 1; 249:74–81. DOI: 10.1007/BF00290238. Available from: https://doi.org/10.1007/BF00290238 [Accessed on: 2026 Jan 24]
32. Bradshaw JE, Hackett CA, Pande B, Waugh R, and Bryan GJ. QTL Mapping of Yield, Agronomic and Quality Traits in Tetraploid Potato (Solanum Tuberosum Subsp. Tuberosum). Theoretical and Applied Genetics. 2008 Jan 1; 116:193–211. DOI: 10.1007/s00122-007-0659-1. Available from: https://doi.org/10.1007/s00122-007-0659-1 [Accessed on: 2026 Mar 13]
33. The Potato Genome Sequencing Consortium. Genome Sequence and Analysis of the Tuber Crop Potato. Nature. 2011 Jul; 475:189–95. DOI: 10.1038/nature10158. Available from: https://www.nature.com/articles/nature10158 [Accessed on: 2024 Dec 29]
34. Hackett CA, McLean K, and Bryan GJ. Linkage Analysis and QTL Mapping Using SNP Dosage Data in a Tetraploid Potato Mapping Population. PLOS ONE. 2013 May 21; 8:e63939. DOI: 10.1371/journal.pone.0063939. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0063939 [Accessed on: 2026 Mar 13]
35. Enciso-Rodriguez F, Douches D, Lopez-Cruz M, Coombs J, and de los Campos G. Genomic Selection for Late Blight and Common Scab Resistance in Tetraploid Potato (Solanum Tuberosum). G3 Genes\textbarGenomes\textbarGenetics. 2018 Jul; 8:2471–81. DOI: 10.1534/g3.118.200273. Available from: https://doi.org/10.1534/g3.118.200273 [Accessed on: 2023 Apr 25]
36. Bourke PM, Voorrips RE, Hackett CA, van Geest G, Willemsen JH, Arens P, Smulders MJM, Visser RGF, and Maliepaard C. Detecting Quantitative Trait Loci and Exploring Chromosomal Pairing in Autopolyploids Using polyqtlR. Bioinformatics. 2021 Nov 5; 37:3822–9. DOI: 10.1093/bioinformatics/btab574. Available from: https://doi.org/10.1093/bioinformatics/btab574 [Accessed on: 2026 Mar 13]
37. Pandey J, Scheuring DC, Koym JW, Endelman JB, and Vales MI. Genomic Selection and Genome-Wide Association Studies in Tetraploid Chipping Potatoes. The Plant Genome. 2022 May; n/a:e20297. DOI: 10.1002/tpg2.20297. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/tpg2.20297 [Accessed on: 2023 Feb 1]

38. Clot CR, Vexler L, de La O Leyva-Perez M, Bourke PM, Engelen CJM, Hutten RCB, van de Belt J, Wijnker E, Milbourne D, Visser RGF, Juranić M, and van Eck HJ. Identification of Two Mutant JASON-RELATED Genes Associated with Unreduced Pollen Production in Potato. Theoretical and Applied Genetics. 2024 Mar 12; 137:79. DOI: 10.1007/s00122-024-04563-7. Available from: https://doi.org/10.1007/s00122-024-04563-7 [Accessed on: 2025 Dec 6]
39. Osborn TC, Chris Pires J, Birchler JA, Auger DL, Jeffery Chen Z, Lee HS, Comai L, Madlung A, Doerge R, Colot V, and Martienssen RA. Understanding Mechanisms of Novel Gene Expression in Polyploids. Trends in Genetics. 2003 Mar 1; 19:141–7. DOI: 10.1016/S0168-9525(03)00015-5. Available from: https://www.sciencedirect.com/science/article/pii/S0168952503000155#FIG1 [Accessed on: 2018 Nov 21]
40. Renny-Byfield S and Wendel JF. Doubling down on Genomes: Polyploidy and Crop Plants. American Journal of Botany. 2014 Oct 1; 101:1711–25. DOI: 10.3732/ajb.1400119
41. Hijmans RJ, Gavrilenko T, Stephenson S, Bamberg J, Salas A, and Spooner DM. Geographical and Environmental Range Expansion through Polyploidy in Wild Potatoes (Solanum Section Petota). Global Ecology and Biogeography. 2007; 16:485–95. JSTOR: 30137903. Available from: https://www.jstor.org/stable/30137903 [Accessed on: 2026 Feb 7]
42. Tai HH, Shannon LM, and Strömvik MV. Polyploidy in Potatoes: Challenges and Possibilities for Climate Resilience. Trends in Genetics. 2025 Aug 1; 41:716–23. DOI: 10.1016/j.tig.2025.03.006. PMID: 40268598. Available from: https://www.cell.com/trends/genetics/abstract/S0168-9525(25)00070-8 [Accessed on: 2026 Feb 7]
43. Bernardo R. Genetics of Breeding Populations. *Breeding for Quantitative Traits in Plants*. Woodbury, Minn. : Stemma Press, 2002. Available from: http://archive.org/details/breedingforquant0000bern [Accessed on: 2025 Oct 25]
44. Lian Q, Tang D, Bai Z, Qi J, Lu F, Huang S, and Zhang C. Acquisition of Deleterious Mutations during Potato Polyploidization. Journal of Integrative Plant Biology. 2019 Jan 1; 61:7–11. DOI: 10.1111/jipb.12748
45. Zhang C, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, and Huang S. The Genetic Basis of Inbreeding Depression in Potato. Nature Genetics. 2019 Mar 14; 51:374–8. DOI: 10.1038/s41588-018-0319-1. Available from: http://www.nature.com/articles/s41588-018-0319-1 [Accessed on: 2019 Mar 20]
46. Song L and Endelman JB. Using Haplotype and QTL Analysis to Fix Favorable Alleles in Diploid Potato Breeding. The Plant Genome. 2023 Jun; 16:e20339. DOI: 10.1002/tpg2.20339. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.20339 [Accessed on: 2025 Sep 13]
47. Smith AB, Cullis BR, and Thompson R. The Analysis of Crop Cultivar Breeding and Evaluation Trials: An Overview of Current Mixed Model Approaches. The Journal of Agricultural Science. 2005 Dec; 143:449–62. DOI: 10.1017/S0021859605005587. Available from: https://www.cambridge.org/core/journals/journal-of-

agricultural-science/article/analysis-of-crop-cultivar-breeding-and-evaluation-trials-an-overview-of-current-mixed-model-approaches/149BDADEE1D04E52616E74A1191A8F69 [Accessed on: 2026 Feb 22]

48. Maris B. Correlations within and between Characters between and within Generations as a Measure for the Early Generation Selection in Potato Breeding. Euphytica. 1988 Apr; 37:205–24. DOI: 10.1007/BF00015117. Available from: https://doi.org/10.1007/BF00015117 [Accessed on: 2024 Nov 21]
49. Gu J, Evers JB, Struik PC, and Driever SM. Architectural Differences between Hybrid Potato Plants Grown from True Seeds and Seedling Tubers Do Not Influence Leaf Photosynthetic Traits. Potato Research. 2025 Oct 17. DOI: 10.1007/s11540-025-09930-5. Available from: https://doi.org/10.1007/s11540-025-09930-5 [Accessed on: 2025 Oct 25]
50. Gopal J. Challenges and Way-forward in Selection of Superior Parents, Crosses and Clones in Potato Breeding. Potato Research. 2015 Jun; 58:165–88. DOI: 10.1007/s11540-015-9292-6. Available from: https://link-springer-com.ezproxy.library.wur.nl/article/10.1007/s11540-015-9292-6 [Accessed on: 2023 Jan 20]
51. Gebhardt C. Bridging the Gap between Genome Analysis and Precision Breeding in Potato. Trends in Genetics. 2013 Apr; 29:248–56. DOI: 10.1016/j.tig.2012.11.006. Available from: https://www.sciencedirect.com/science/article/pii/S0168952512001904 [Accessed on: 2023 Feb 1]
52. Jansky SH and Peloquin SJ. Advantages of Wild Diploid Solanum Species over Cultivated Diploid Relatives in Potato Breeding Programs. Genetic Resources and Crop Evolution. 2006 Jun 21; 53:669–74. DOI: 10.1007/s10722-004-2949-7. Available from: https://link.springer.com/article/10.1007/s10722-004-2949-7 [Accessed on: 2021 Jan 29]
53. Tai G. ESTIMATION OF GENERAL AND SPECIFIC COMBINING ABILITIES IN POTATO. Canadian Journal of Genetics and Cytology. 1976 Sep 1; 18:463–70. DOI: 10.1139/g76-056. Available from: http://www.nrcresearchpress.com/doi/10.1139/g76-056 [Accessed on: 2025 May 17]
54. Lemaga B and Caesar K. Relationships between Numbers of Main Stems and Yield Components of Potato (Solanum Tuberosum L. Cv. Erntestolz) as Influenced by Different Daylengths. Potato Research. 1990 Jun; 33:257–67. DOI: 10.1007/BF02358455. Available from: https://link.springer.com/article/10.1007/BF02358455 [Accessed on: 2021 Jun 28]
55. Haynes KG. Variance Components for Yield and Specific Gravity in a Diploid Potato Population after Two Cycles of Recurrent Selection. 2001 :69–75
56. Hirsch CN, Hirsch CD, Felcher K, Coombs J, Zarka D, Van Deynze A, De Jong W, Veilleux RE, Jansky S, Bethke P, Douches DS, and Buell CR. Retrospective View of North American Potato (Solanum Tuberosum L.) Breeding in the 20th and 21st Centuries. G3: Genes, Genomes, Genetics. 2013; 3:1003–13. DOI: 10.1534/g3.113.005595
57. Bethke PC, Halterman DA, and Jansky SH. Potato Germplasm Enhancement Enters the Genomics Era. Agronomy. 2019 Oct; 9:575. DOI: 10.3390/agronomy9100575.

Available from: https://www.mdpi.com/2073-4395/9/10/575 [Accessed on: 2026 Mar 7]

58. Lindhout P, Meijer D, Schotte T, Hutten RC, Visser RG, and Eck HJ van. Towards F1 hybrid seed potato breeding. Potato Research 2011; 54:301–12
59. Bradshaw JE. A Brief History of the Impact of Potato Genetics on the Breeding of Tetraploid Potato Cultivars for Tuber Propagation. Potato Research. 2022 Sep; 65:461–501. DOI: 10.1007/s11540-021-09517-w. Available from: https://link.springer.com/article/10.1007/s11540-021-09517-w [Accessed on: 2023 May 4]
60. Crow JF. 90 Years Ago: The Beginning of Hybrid Maize. Genetics. 1998; 148:923–8. Available from: https://academic.oup.com/genetics/article-abstract/148/3/923/6034434 [Accessed on: 2025 Aug 9]
61. Cowling WA, Gaynor RC, Antolín R, Gorjanc G, Edwards SM, Powell O, and Hickey JM. In Silico Simulation of Future Hybrid Performance to Evaluate Heterotic Pool Formation in a Self-Pollinating Crop. Scientific Reports. 2020 Mar; 10:4037. DOI: 10.1038/s41598-020-61031-0. Available from: https://www.nature.com/articles/s41598-020-61031-0 [Accessed on: 2023 May 6]
62. Labroo MR, Studer AJ, and Rutkoski JE. Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. Frontiers in Genetics. 2021 Feb 24; 12:234. DOI: 10.3389/FGENE.2021.643761
63. Almekinders CJM, Chilver AS, and Renia HM. Current Status of the TPS Technology in the World. Potato Research. 1996 Jun 1; 39:289–303. DOI: 10.1007/BF02360921. Available from: https://doi.org/10.1007/BF02360921 [Accessed on: 2026 Mar 13]
64. Stockem JE, Korontzis G, Wilson SE, Vries ME de, Eeuwijk FA, and Struik PC. Optimal Plot Dimensions for Performance Testingfof Hybrid Potato in the Field. Potato Research. 2021
65. Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, Würschum T, Mock HP, Matros A, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Gowda M, Longin CFH, and Reif JC. Genome-Based Establishment of a High-Yielding Heterotic Pattern for Hybrid Wheat Breeding. Proceedings of the National Academy of Sciences of the United States of America. 2015 Dec 22; 112:15624–9. DOI: 10.1073/pnas.1514547112. PMID: 26663911. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26663911 [Accessed on: 2019 May 3]
66. Peixoto MA, Coelho IF, Leach KA, Bhering LL, and Resende Jr. MFR. Simulation-Based Decision-Making and Implementation of Tools in Hybrid Crop Breeding Pipelines. Crop Science. 2024; 64:110–25. DOI: 10.1002/csc2.21139. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/csc2.21139 [Accessed on: 2025 Aug 16]
67. Hougas R and Peloquin S. The Potential of Potato Haploids in Breeding and Genetic Research. American Potato Journal. 1958; 35:701–7
68. Chase S. Analytic Breeding in SOLANUM TUBEROSUM L. A Scheme Utilizing Parthenotes and Other Diploid Stocks. Canadian Journal of Genetics and Cytology. 1963; 5:359–63

69. Uijtewaal BA, Huigen DJ, and Hermsen JGTh. Production of Potato Monohaploids (2n=x=12) through Prickle Pollination. Theoretical and Applied Genetics. 1987 Sep 1; 73:751–8. DOI: 10.1007/BF00260786. Available from: https://doi.org/10.1007/BF00260786 [Accessed on: 2025 Aug 2]
70. Hosaka K and Hanneman RE. Genetics of Self-Compatibility in a Self-Incompatible Wild Diploid Potato Species Solanum Chacoense. 2. Localization of an S Locus Inhibitor (Sli) Gene on the Potato Genome Using DNA Markers *. Euphytica. 1998; 103:265–71
71. Jing X, Qian D, Jiang X, Wang P, Bao H, Li D, Zhu Y, Zhu G, and Zhang C. Genetic Diversity of the Self-Incompatibility Locus in Diploid Potato. Journal of Integrative Agriculture. 2025 Apr 1; 24:1448–60. DOI: 10.1016/j.jia.2024.12.011. Available from: https://www.sciencedirect.com/science/article/pii/S2095311924004209 [Accessed on: 2026 Mar 5]
72. Ma L, Zhang C, Zhang B, Tang F, Li F, Liao Q, Tang D, Peng Z, Jia Y, Gao M, Guo H, Zhang J, Luo X, Yang H, Gao D, Lucas WJ, Li C, Huang S, and Shang Y. A nonS-locus F-box Gene Breaks Self-Incompatibility in Diploid Potatoes. Nature Communications. 2021 Jul; 12:4142. DOI: 10.1038/s41467-021-24266-7. Available from: https://www.nature.com/articles/s41467-021-24266-7 [Accessed on: 2023 Feb 1]
73. Eggers EJ, van der Burgt A, van Heusden SAW, de Vries ME, Visser RGF, Bachem CWB, and Lindhout P. Neofunctionalisation of the Sli Gene Leads to Self-Compatibility and Facilitates Precision Breeding in Potato. Nature Communications. 2021 Jul; 12:4141. DOI: 10.1038/s41467-021-24267-6. Available from: https://www.nature.com/articles/s41467-021-24267-6 [Accessed on: 2023 Jan 20]
74. Pallais N. True Potato Seed Quality. Theoretical and Applied Genetics. 1987 Apr 1; 73:784–92. DOI: 10.1007/BF00289380. Available from: https://doi.org/10.1007/BF00289380 [Accessed on: 2026 Mar 7]
75. Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, and Huang S. Genome Design of Hybrid Potato. Cell. 2021 Jul 22; 184:3873–3883.e12. DOI: 10.1016/J.CELL.2021.06.006
76. De Vries ME, Adams JR, Eggers Ej, Ying S, Stockem JE, Kacheyo OC, Van Dijk LCM, Khera P, Bachem CW, Lindhout P, and Van Der Vossen EAG. Converting Hybrid Potato Breeding Science into Practice. Plants. 2023 Jan; 12:230. DOI: 10.3390/plants12020230. Available from: https://www.mdpi.com/2223-7747/12/2/230 [Accessed on: 2024 Jun 3]
77. Villanueva B, Wray NR, and Thompson R. Prediction of Asymptotic Rates of Response from Selection on Multiple Traits Using Univariate and Multivariate Best Linear Unbiased Predictors. Animal Science. 1993 Aug; 57:1–13. DOI: 10.1017/S0003356100006541. Available from: https://www.cambridge.org/core/journals/animal-science/article/prediction-of-asymptotic-rates-of-response-from-selection-on-multiple-traits-using-univariate-and-multivariate-best-linear-unbiased-predictors/DF8BD07F1E3C799CD6F0C9F43CFAB55D [Accessed on: 2025 May 30]

78. Endelman JB, Carley CA, Bethke PC, Coombs JJ, Clough ME, da Silva WL, De Jong WS, Douches DS, Frederick CM, Haynes KG, Holm DG, Miller JC, Muñoz PR, Navarro FM, Novy RG, Palta JP, Porter GA, Rak KT, Sathuvalli VR, Thompson AL, and Yencho GC. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics. 2018 May 1; 209:77–87. DOI: 10.1534/genetics.118.300685. PMID: 29514860. Available from: https://pubmed.ncbi.nlm.nih.gov/29514860/ [Accessed on: 2021 Jan 29]

2

Little heterosis found in Diploid Hybrid Potato — The genetic underpinnings of a new hybrid crop

This chapter is based on:

Adams JR et al. Little Heterosis Found in Diploid Hybrid Potato: The Genetic Underpinnings of a New Hybrid Crop. G3: Genes, Genomes, Genetics. 2022 Jun; 12. doi: 10.1093/g3journal/jkac076. PMID: 35460241

Abstract

Hybrid potato breeding has become a novel alternative to conventional potato breeding allowing breeders to overcome intractable barriers (e.g. tetrasomic inheritance, masked deleterious alleles, obligate clonal propagation) with the benefit of seed-based propagule, flexible population design, and the potential of hybrid vigour. Until now, however, no formal inquiry has adequately examined the relevant genetic components for complex traits in hybrid potato populations. In this present study, we use a two-step multivariate modelling approach to estimate the variance components to assess the magnitude of the general and specific combining abilities (GCA and SCA, respectively) in diploid hybrid potato (DHP). SCA effects were identified for all yield components studied here warranting evidence of non-additive genetic effects in hybrid potato yield. However, the estimated GCA effects were on average two times larger than their respective SCA quantile across all yield phenotypes. Tuber number GCA's and SCA's were found to be highly correlated with total yield's genetic components. Tuber volume was shown to have the largest proportion of additive and non-additive genetic variation suggesting under-selection of this phenotype in this population. The prominence of additive effects found for all traits presents evidence that the mid-parent value alone is useful for hybrid potato evaluation. Heterotic vigour stands to be useful in bolstering simpler traits but this will be dependent on target phenotypes and market requirements. This study represents the first diallel analysis of its kind in diploid potato using material derived from a commercial hybrid breeding programme.

2.1 Key Message

Hybrid vigour was detected for multiple traits in diploid hybrid potato. Additive gene action was most prominent in tuber yield and should be the primary target within hybrid breeding programmes.

2.2 Introduction

Potato (*Solanum tuberosum*), a plant species once isolated to the continents of southern and central America, is now a crop that spans over 17 million hectares of crop-land worldwide[2]. It is the most prominent of non-cereals and is considered by many a major keystone in guaranteeing food security for both local and global communities. Prized for their edible starch-rich tubers, potato meets the demand of several key industries including the fresh, processing, and seed potato markets with a global gross value of 140.5 billion USD as of 2019. As a field crop, potato has a competitive harvest index of 0.85 (in contrast to 0.4-0.6 seen in other crops) in conjunction with a high water productivity[3,4]. There is also ample variation in potato's tuberization timing requiring as few as 75 days from planting to harvest. All the above make potato a highly productive crop amenable to a variety of cropping systems capable of supplying valuable starch with less agronomic input.

Despite potato's growing economic and societal importance, rates of crop improvement in complex traits have not kept in step with other major crops over the past century[5,6]. Reasons for these deficits in genetic gain are numerous (e.g. market segmentation, large inventory of quality traits, etc.) but many of them stem from the complexities of potato's evolution and domestication. Potato's tetraploidy is an oft-cited stumbling block for breeders impeding the ability to fix beneficial loci, and conversely, remove deleterious sites harboured across the genome[7,8]. Not only does polyploidy mask deleterious loci from traditional forces of selection, but it also impacts the length of time for site fixation even under genetic drift leading to greater maintenance of heterozygosity over time[9]. Taken together with a very strong self-incompatibility mechanism, potato could best be described as a fortified heterozygous out-crosser. These biological realities shaped potato breeding from the beginning with breeders conducting crosses between promising heterozygous individuals followed by the evaluation of large nurseries in search for decent complementation[10]. These F1 nurseries were then subjected to as many as eight subsequent rounds of clonal selection until only elite candidates were left[11]. While in some ways this method of clonal breeding is quite efficient (all genetic factors are effectively *fixed* at the creation of the F1), it is widely known for being a long process from generation of the nursery to variety release. Because the success of clonal breeding is highly dependent on the generation of enough novel genotypes in the F1, it takes as many as 9 years to sufficiently bulk tubers in conjunction with applying appropriate selection pressure[12,13]. It should be noted that while there have been proposals to optimise conventional clonal breeding[14,15], many of the aforementioned issues are simply implicit to breeding tetraploid potato.

One solution to this comprehensive set of challenges is the adaptation of potato from a tetraploid clonal crop to that of a diploid inbred-hybrid one, an idea which has existed in some form for over 60 years[16]. The benefits of such a change, if possible, are manifold; Diploids only take one generation to half their heterozygosity in contrast to an autotetraploid which takes upwards of four generations making the production of pure-breeding lines plausible in the former. As an extension of this, superior genetic performance in the final marketed variety is not dependent on a single crossing event that generated the

original F1 (as it is in conventional clonal breeding) but is accomplished through multiple stages (e.g. parental pool improvement, parental line development, hybrid crossing, etc.). This is not to mention other logistical niceties such as the ease of producing & storing true potato seed over vegetative propagule[17–19]. Despite the potential of diploid potato, however, it was not until the cusp of the 21st century that it became broadly feasible. Many picked up on the work of[20] and began the process of generating self compatible populations through the use of *Sli*. (Lindhout et al., 2011) was one of the first to confirm the commercial potential generating diploid potato populations capable of inbreeding using a *Sli* donor. Several subsequent studies not only corroborated that inbred populations in diploid potato were possible[21], but hybrids generated from these populations resulted in a crop that could compete in the same space as tetraploid potato[22,23]. While diploid hybrid potato (DHP) populations are now extant across the world, there is at this time little known about the genetic components controlling complex traits as DHP is still a young hybrid crop. Understanding this is an imperative for potato breeders in order to structure breeding programmes which are able to best exploit the genetic variation available to DHP.

We set out to inspect tuber yield in a large DHP test-cross. To do this, we performed a joint evaluation of total yield along with two of its simpler yield components, average tuber volume and total tuber number, and partitioned their underlying genetic effects into additive and non-additive components. This trait and genetic decomposition was done to inspect two broad questions: (1) Which tuber phenotypes in this population were responsible for the variation seen in total yield and (2) are these yield phenotypes primarily under the control of additive or non-additive gene action? This latter question holds particular weight as it gives insight on where the focal point of DHP breeding should lie. We put forward a two-part modelling approach to utilise intra-block information to estimate the general and specific combining abilities of our hybrid parents and crosses, respectively. Our study presents the first diallel study in DHP using highly-inbred parents derived from a commercial breeding programme.

2.3 Materials and methods

Crosses and Trials

A panel of 400 inbred parents were selected and crossed according to distinct selection criteria related to fertility and agronomic traits yielding 806 successful F1 crosses. These parents were produced from an experimental population derived from several backgrounds including *tuberosum* and several wild species (e.g. *Solanum chacoense*)[24]. In the Spring of 2019, all hybrid TPS were sown in trays and grown out in a greenhouse. In May, all seedlings were transplanted at stage 105 development (See[25]) into two field trials located in the Dutch towns of Est and Heelsum. Both trials utilised a double ridge design with eight plants per ridge with a total of sixteen plants per plot; this design was chosen to minimize within-plot variation while reducing planting costs across each trial[26]. Plots were organised in an augmented randomised complete block design with two blocks and three internal controls used across each block. All 806 F1 hybrids were planted in Heelsum

Table 2.1: Agronomic properties of the screening trials conducted in Est and Heelsum

					Soil Composition[a]			
Locations	Year	Rows	Columns	Hybrids	Sand	Silt	Clay	Organic matter
Est	2019	20	65	608	23	45	28	3
Heelsum	2019	20	85	806	76	16	2	6

[a] Soil characteristics presented as percentage

with a subset of 608 hybrids planted in Est. Trial conditions were similar with regards to field management and scoring. One distinguishing factor between trials were their soil conditions with Est being characterised by a light clay composition and Heelsum conversely by distinctively sandy conditions (see Table 2.1). Both trials were conducted through the summer until haulm killing in early September followed by subsequent harvest two weeks later. All hybrids were scored by several criteria including relevant yield related traits which are our primary focus for this study, i.e., total yield, tuber number and tuber volume. Total tuber number (TN) was measured as the total number of tubers harvested from a given plot of sixteen plants. Tuber volume (TV) was calculated using an average over all tubers harvested per plot using a tuber's length, width, and depth dimensions to calculate volume using an ellipsoid approximation. Lastly, total yield (TY) was calculated through a transformation of the total tuber weight of a plot to estimate the approximate yield in units of $Mg \cdot Ha^{-1}$. These traits were collected for all tubers above 20 mm in length via an automated pipeline described in[22].

Spatial Models

This present study used a two-step modelling approach where each field trial was modelled separately accounting for factors like field design, control effects, and spatial heterogeneity allowing for extraction of spatial trends and de-trended phenotype data. This was followed by modelling of genetic components simultaneously for all phenotypes. The first step was accomplished by partitioning field effects into local and global trends using two-dimensional penalised splines. This was performed using the Spatial Analysis of Field Trials with Splines (SpATS) library available through the comprehensive R archive network (CRAN)[27]. While many attested methods capable of handling geo-spatial trends exist, the spatial smoothing approach offered through SpATS was chosen for a few reasons. Often, genetic modelling requires the creation of many spatial models with different spatial structures in order to identify the most satisfactory spatial model. SpATS, conversely, does not follow this procedure and is capable of offering comparable genotype estimates with the best traditional spatial model[28]. Along with this, SpATS provides a number of internal methods allowing for intelligible and simple model diagnostics to help elucidate the predominate factors for a given field trial. We chose to model field dimensions using SpATS' PS-ANOVA method which is capable of taking the bivariate surface and decomposing it into multiple spatial components all defined by one smoothing parameter[29]. The resultant model equation is then:

$$y_{chmn} = \beta_c + g_h + r_m + c_n + f_{ps}(m, n) + \varepsilon_{chmn} \tag{2.1}$$

where β_c is a fixed effect for whether the hybrid was a control, g_h is a random effect for the h^{th} hybrid, m and n are random effects for row m and column n. The row and column co-ordinates were also used within f_{ps} which represents the two-dimensional penalised-spline function. The PS-ANOVA was parametrised using 19 & 83 internal knots for Heelsum and 19 & 63 internal knots for Est. The large number of internal knots resulted in longer computational time, but were selected to mirror the number of plots along each row and column for each trial. Third degree polynomial B-splines with second degree penalties were used for all spatial models, from which, spatial trends were derived and then subsequently used to de-trend the phenotype data for each trait:

$$y^*_{hk} = y_{hmn} - (r_m + c_n + f_{ps}(m, n)) \tag{2.2}$$

where y^* represents the corrected phenotype with systematic spatial trends removed. Each spatial trend was presented as a percentage deviation from the trial mean (see Figures 2.1 & 2.2). Along with this, every spatial model was evaluated on the basis of effective and nominal dimension number estimated for each model effect. These were used to evaluate the number of parameters estimated for smoothing and random terms (See table's S1 & S2). Taking the ratio between effective and nominal dimensions for random hybrid effects have the benefit of being interpreted as a generalised heritability where the effective dimension number for a hybrid genotype effect (the trace of its hat matrix) is divided by its nominal dimension (the rank of its design matrix) allowing for a direct assessment of genetic variation exhibited within a field trial[30].

Genetic models

For genetic modelling, F1 hybrids were included in subsequent analysis based upon two criteria: their presence in both screening trials, and whether both parents of a hybrid were utilised in at least two crosses. The former criteria was to ensure estimation of each genotype location combination while the latter was to exclude unconnected crossing sets to guarantee demarcation of parental and cross-wise effects. This resulted in the selection of 225 parental lines which gave rise to 495 F1 hybrid progeny. This panel of hybrids was first utilised in the following multi-trait multi-location model:

$$y^*_{dfgk} = \mu_d + \beta_{df} + h_{dg} + t_{dfg} + \varepsilon_{dfgk} \tag{2.3}$$

where each trait, d, is treated as having its own mean (μ_d) and field effect (β_{df}) for trial location, f. Hybrid effects are specified as h_{dg} for trait, d, and hybrid, g, while t_{dfg} is the hybrid by trial by trait interaction for hybrid, g, trial, f, and trait, d. ε_{dfgk} is the residual for trait, d, hybrid, g, field trial, f, and replicate, k. All random effects were modelled as being drawn from a multivariate normal distribution centred around zero and an unstructured covariance matrix for each effect. This includes the hybrid effect ($h \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_h}\sigma^2_h)$), hybrid by field trial interaction effect ($t \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_t}\sigma^2_t)$),

and the residual ($\varepsilon \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_n}\sigma^2_\varepsilon)$). From this hybrid model, best linear unbiased predictions (BLUPs) were made for each phenotype and hybrid over all trials ($E[y_{dg}|\mathbf{h}]$) as well as conditioned on each trial ($E[y_{dfg}|\mathbf{h}, \mathbf{t}]$) (See Figure 2.3); variance components (excluding covariances) were also extracted from this hybrid model for all three tuber phenotypes (Table S3).

The intent of our paper is not merely to retrieve hybrid estimates, but further decompose these estimates into distinct additive and non-additive components. In the context of plant populations, this is traditionally done through a series of controlled crosses between a set of parents which allows for the separation of the parental mean, or the general combining ability (GCA), and the deviation from the expected mean of a cross, or the specific combining ability (SCA)[31]. These two parameters also have the benefit of being interpreted in terms of genetic variances of a population. The variance attributable to GCA is equal to the covariance between half-siblings while the variance attributed to SCA is equal to the covariance between to full-siblings subtracted by twice covariance of half-siblings[32]. Such models have been made for a variety of population designs (full diallel, half diallel, factorial, etc.) with different effect structures depending on the intent of inquiry. For example, genetic effects can be modelled as fixed if the interest is to provide valid performance estimates for a given cross or they can be treated as random if the variance of effect sampled from a population is desired to be studied[33]. Additionally, these models can be expanded or simplified accommodating reciprocal effects, trial location or environment interactions, population structures, and so forth. For our purposes, we extend Griffing's model II[34] into a multivariate context where hybrid yield can be described as following:

$$y^*_{dfijk} = \mu_d + \beta_{df} + g_{di} + g_{dj} + t_{dfi} + t_{dfj} + \varepsilon_{dfijk} \tag{2.4}$$

where, identical to equation (2.3), μ_d and β_{df} are fixed effects for a global mean and field trial effect, respectively, for trait, d, and field trial, f. g_{di} and g_{dj} are the GCAs of parents' i and j, respectively, with t_{dfi} and t_{dfj} being their respective field trial and parental interactions for trait, d, and trial location, f. ε_{dfijk} is the model's residual for replicate, k, on progeny of parents', i and j, evaluated in trial, f, for trait, d. We assume all random effects to be drawn from a multivariate normal distribution, that is, the maternal and paternal GCAs ($g \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_g}\sigma^2_{gca})$), GCA by environment interactions ($t \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_t}\sigma^2_{gxe})$), and the residual effects ($\varepsilon \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I}_\varepsilon\sigma^2_\varepsilon)$) with zero covariance between these random effects. This model, containing only the additive genetic effects, will hereon be denoted as M_0.

This model can be expanded further to include hybrid cross-wise effects with the addition of the SCA and SCA by environment interaction effects. This final model then has the form:

$$y^*_{dfijk} = \mu_d + \beta_{df} + g_{di} + g_{dj} + t_{dfi} + t_{dfj} + s_{dij} + r_{dfij} + \varepsilon_{dfijk} \tag{2.5}$$

where s_{dij} is the SCA of parents' i and j with r_{dfij} being their respective interaction

with trial location, f, all for trait, d. Distributional properties of random effects follow (2.4) with the addition of two random effects, SCA ($s \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_s}\sigma^2_{sca})$) and SCA by environment interaction ($r \sim \mathsf{MVN}(\mathbf{0}, \Sigma \otimes \mathsf{I_r}\sigma^2_{sxe})$) and will be denoted as M_f going forward. Because the underlying crossing sets are sparse, identifiability of models M_0 and M_f was tested following[35] to ensure that statistically valid estimates could be derived from all genetic models.

Variance Ratios & Genetic Correlations

To study all relevant effects, variance components were estimated and extracted from models M_0 and M_f. These variance components were used in two general forms: (1) to derive ratios of effects within traits and (2) produce genetic correlations between traits and trial locations. These components were first used to derive several important genetic parameters including variation due to additive genetic effects ($2 \cdot \mathsf{Diag}(\Sigma\sigma^2_{gca}) = (\sigma^2_{a_{TY}}, \sigma^2_{a_{TN}}, \sigma^2_{a_{TV}})^T$), variation due to dominance ($\mathsf{Diag}(\Sigma\sigma^2_{sca}) = (\sigma^2_{d_{TY}}, \sigma^2_{d_{TN}}, \sigma^2_{d_{TV}})^T$) and their respective environmental interactions ($2 \cdot \mathsf{Diag}(\Sigma\sigma^2_{gxe}) = (\sigma^2_{ae_{TY}}, \sigma^2_{ae_{TN}}, \sigma^2_{ae_{TV}})^T$ & $\mathsf{Diag}(\Sigma\sigma^2_{sxe}) = (\sigma^2_{de_{TY}}, \sigma^2_{de_{TN}}, \sigma^2_{de_{TV}})^T$). Proportion of total phenotypic variance was then examined with respect to these genetic variances along with each trait's residual variance (Figure 2.4). These genetic parameters were then used to calculate several variance ratios (See Table 2.2) including broad and narrow-sense heritabilities, H^2 ($\frac{\sigma^2_a+\sigma^2_d}{\sigma^2_p}$) and h^2 ($\frac{\sigma^2_a}{\sigma^2_p}$), respectively, dominance ratios ($d^2 = \frac{\sigma^2_d}{\sigma^2_p}$), additive portion of genetic variation ($\frac{\sigma^2_a}{\sigma^2_G}$), and additive by trial location portion of total genetic by environment variation ($\frac{\sigma^2_{ae}}{\sigma^2_{GE}}$). Total phenotypic variation (σ^2_p) was computed by scaling σ^2_{ae} and σ^2_{de} by the total number of field trials and σ^2_ε by the product of the number of field trials and the number of replicates used within each trial. Thus, $\sigma^2_p = \sigma^2_a + \frac{\sigma^2_{ae}}{2} + \sigma^2_d + \frac{\sigma^2_{de}}{2} + \frac{\sigma^2_\varepsilon}{4}$.

Second, variance components from model M_f were used to estimate genetic correlations in GCA and SCA effects. These include genetic correlations between traits (e.g. $\frac{\mathsf{cov}(\mathsf{gca_{TY}},\mathsf{gca_{TV}})}{\sigma_{\mathsf{TY}}\sigma_{\mathsf{TV}}}$), intra-class correlation coefficients between trial locations (e.g. $\frac{\sigma^2_{\mathsf{gca}}}{\sigma^2_{\mathsf{gca}}+\sigma^2_{\mathsf{gxe}}}$), and genetic correlations between traits and trial locations (e.g. $\frac{\mathsf{cov}(\mathsf{gca_{TY}},\mathsf{gca_{TV}})+\mathsf{cov}(\mathsf{gxe_{TY}},\mathsf{gxe_{TV}})}{\sqrt{(\sigma^2_{\mathsf{gca_{TY}}}+\sigma^2_{\mathsf{gxe_{TY}}})\cdot(\sigma^2_{\mathsf{gca_{TV}}}+\sigma^2_{\mathsf{gxe_{TV}}})}}$). These were computed for the GCA and GCA by trial location effects (Figure 2.5) as well as SCA and SCA by trial location effects (Figure 2.6). Each of these are presented with multivariate scatterplots and marginal BLUP distributions for all effects.

Hypothesis testing

To evaluate statistical evidence of heterosis (through SCA term), we perform a hypothesis testing procedure on the original M_0 (equation (2.4)) and M_f (equation (2.5)) models as well as on the univariate analogues for each trait. This was done to assess the meaningful addition of SCA effects with respect to each phenotype without consideration of extra covariance parameters in M_0 and M_f. M_{0d}, that is, the univariate form of model (2.4) for trait, d, represents a null model where the additional effects from M_{fd}, the univariate form of model (2.5) for trait, d, are constrained to zero. Therefore, we can construct a

non-standard hypothesis test where:

$$\begin{aligned} H_0 &: \sigma^2_{sca}, \sigma^2_{sxe} = 0 \\ H_1 &: \sigma^2_{sca}, \sigma^2_{sxe} > 0 \end{aligned} \tag{2.6}$$

which can be evaluated directly through the following likelihood ratio test where:

$$\Lambda(y_d) = -2\ln\left(\frac{\mathcal{L}(M_{0d})}{\mathcal{L}(M_{fd})}\right) = 2(loglik(M_{fd}) - loglik(M_{0d})) \tag{2.7}$$

This test is non-standard because it follows a special case where testing is occurring on the boundary of the parameter space which is often taken into account using a mixed χ^2 distribution following[36]. For our testing purposes, we used a $0.25\chi^2_0 + 0.50\chi^2_1 + 0.25\chi^2_2$ mixture distribution. These tests are also accompanied with the Akaike information criterion (AIC) for each univariate and multivariate set of models (Table 2.3). Under H_1, the hybrid genetic effect from our original model (2.3) should equal the sum of the GCA & SCA effects specified in model (2.5). Our modelling procedure for all multivariate models began with the estimation of variance components and production of genetic correlations on BLUP's in the univariate analogues (i.e. M_{0d} and M_{fd}) which were used to initialise the unstructured covariance matrices for all random effects in model's M_0 and M_f. All models were fitted through restricted (residual) maximum likelihood (REML) using ASReml-R 4[37]. REML based procedures have come into popular usage over the past two decades due to their ability to provide estimators both consistent and asymptotically normal even under conditions with non-orthogonal sets of random predictors which is particularly useful while using sparse crossing designs[38]. This is not to mention the volume of diallel-based literature where REML is the invoked method of choice for reasons which will not be discussed here ([39] provide an excellent review on the topic).

2.4 Results

Spatial Components

Spatial models for the Est and Heelsum trials were estimated for each yield component. All spatial models successfully converged with credible spatial trends for both trial locations. Model residuals for all environmental models showed little to no evidence of deviation from normality; the following suggest successful delineation of spatial components for all traits in both field trials.

Total yield and tuber number exhibited evidence for strong local trends with a row effect contributing to the spatial trend in Est (Figure 2.1). While these same components also impacted tuber volume, it was not nearly so prominent (See S1). The similar magnitude of row effects on tuber number and total yield can also be observed through the ratio of effective and nominal dimensions which were identical for these two traits (0.68) in contrast to tuber volume (0.47). Additionally, the effective nominal dimension ratio for

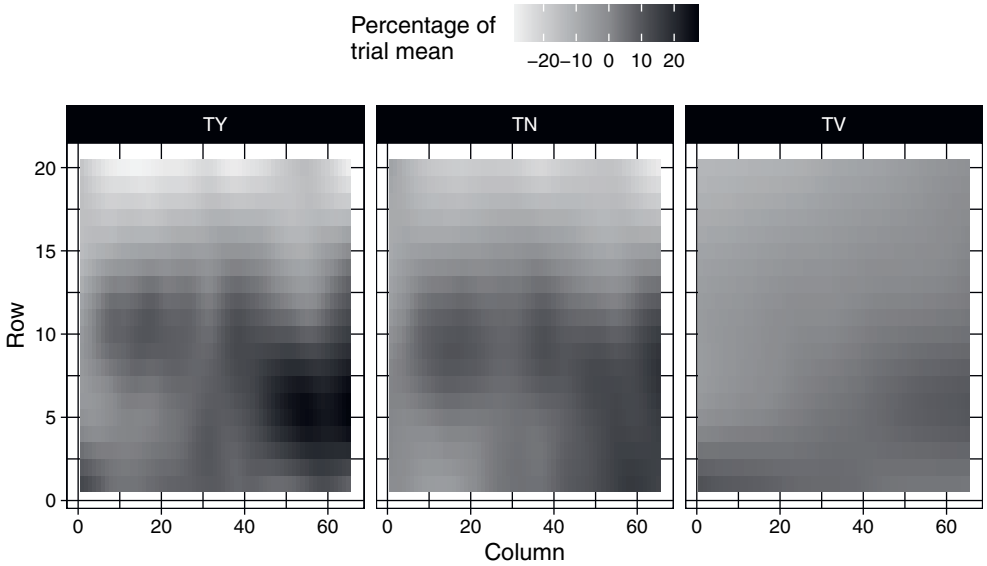


Figure 2.1: Estimated spatial trends scaled by trial mean for total yield (TY; $\frac{Mg}{Ha}$) , tuber number (TN; Number of tubers per plot), and tuber volume (TV; Average cm^3 per plot) within the Est screening trial.

hybrids (i.e. a generalised heritability) was highest in tuber volume followed by total yield and lastly tuber number (See Table S1).

Overall, the spatial trends in the Heelsum trial were less severe than those evaluated in the screening trial in Est. Interquartile ranges of field effects were between -8.96% and 9.04% of the trial mean for total yield in Est while the range was -2.85% and 3.55% of the trial mean for yield in Heelsum (Figure 2.2). The differences in magnitude of spatial components between these two trials were also similar for tuber number and tuber volume. Nonetheless, the estimated spatial components did have a modest effect on all yield related components in Heelsum. In particular, random effects for the field column had a minor impact on tuber volume (0.55), tuber number (0.48), and total yield (0.45) (See Table S2). General heritability estimates for Heelsum were akin to those observed in Est with values of 0.9, 0.82, and 0.88 for tuber volume, tuber number, and total yield, respectively.

Hybrid estimates

Using spatially corrected phenotypes from model (2.1), we extracted both marginal and conditioned hybrid BLUP's from model (2.3) (Figure 2.3). Generally, hybrid performance was far greater in Heelsum over Est. The trial mean for total yield was 13.1 and 22.4 $Mg \cdot Ha^{-1}$ for Est and Heelsum, respectively. Likewise, average tuber number was 136 tubers in Est and 199 tubers in Heelsum. Trial means for tuber volume were comparatively more stable between trial locations with means of 21.3 cm^3 in Est and 24.2 cm^3 in Heelsum. Along with differences in mean hybrid performance, there was also greater dispersion of phenotypes in Heelsum than in Est. This was especially apparent for total yield in Heelsum

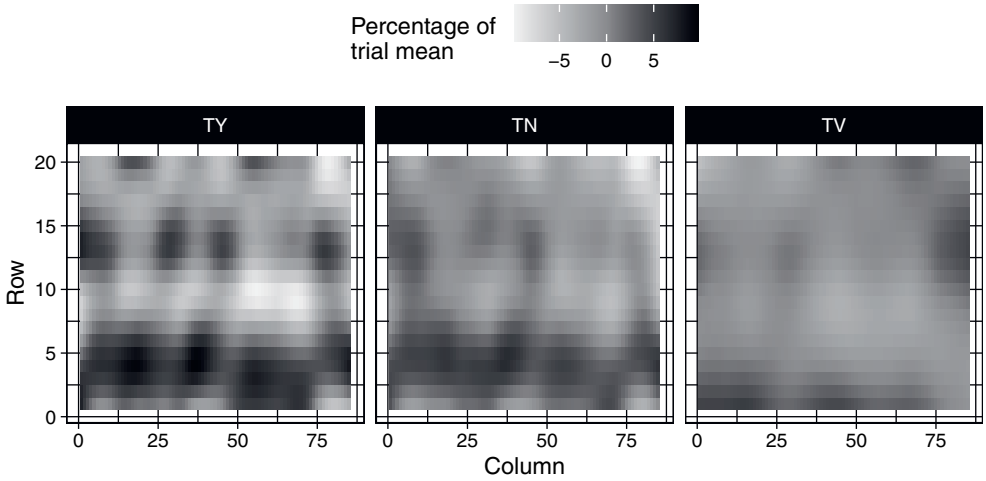


Figure 2.2: Estimated spatial trends scaled by their trial mean for total yield (TY; $\frac{Mg}{Ha}$) , tuber number (TN; Number of tubers per plot), and tuber volume (TV; Average cm^3 per plot) within the Heelsum screening trial.

which displayed a standard error twice the size of total yield in Est. This same marked difference could also be seen in tuber number where BLUP's in Heelsum exhibited a standard error 1.6 times greater than that which was found in Est. Tuber volume in contrast to the other phenotypes exhibited similar BLUP distributions between both trial locations. Examining these BLUP's in light of each trait's variance components (See Table S3) suggests that tuber volume showed the greatest stability of all three phenotypes.

Variance Ratios

Variance estimates were derived for all specified random effects for models M_0 and M_f. Tuber volume not only exhibited the largest proportion of variance explained by SCA ($d^2 = 0.14$), but also had the largest total genetic variance of any trait ($H^2 = 0.86$) (Figure 2.2). Tuber number and total yield harboured similar proportion's of SCA ($d^2 = 0.12$) with a considerable portion of non-additive effects being partitioned in the SCA by environment interaction (Figure 2.4). Broad-sense heritabilities were quite similar between tuber number (0.74) and total yield (0.71) with the primary difference between the two traits being the partitioning of genotype by environment interactions ($\frac{\sigma^2_{ae}}{\sigma^2_{GE}}$ equal to 0.78 in total yield in contrast to 0.65 in tuber number) and size of the residual variance (Figure 2.4). The additive genetic component was the largest genetic effect across all traits with the ratio of additive genetic variance being identical in total yield and tuber number (0.84) and nearly identical in tuber volume (0.83). Between models M_0 and M_f, partitioning of variance changes most drastically for σ^2_{ae}. These were much larger in M_f in tuber number and total yield with the incorporation of the SCA main effect and SCA by environment interaction (Figure 2.4).

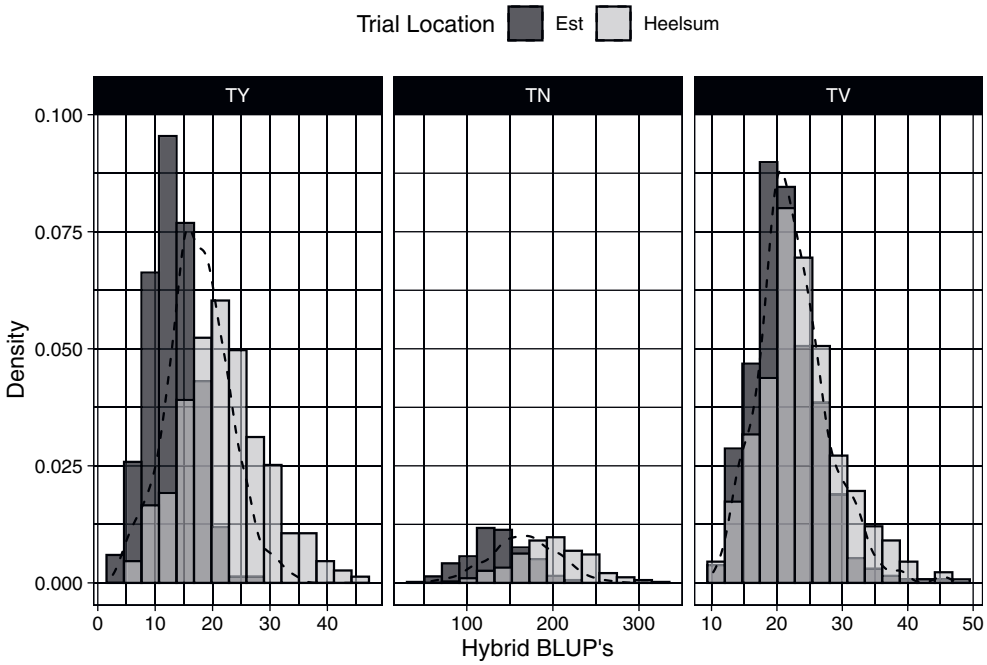


Figure 2.3: Hybrid best linear unbiased predictions for total yield (TY; $\frac{Mg}{Ha}$) , tuber number (TN; Number of tubers per plot), and tuber volume (TV; Average cm^3 per plot) conditioned on Est and Heelsum trials. Across-location BLUP's visualised through black density curve.

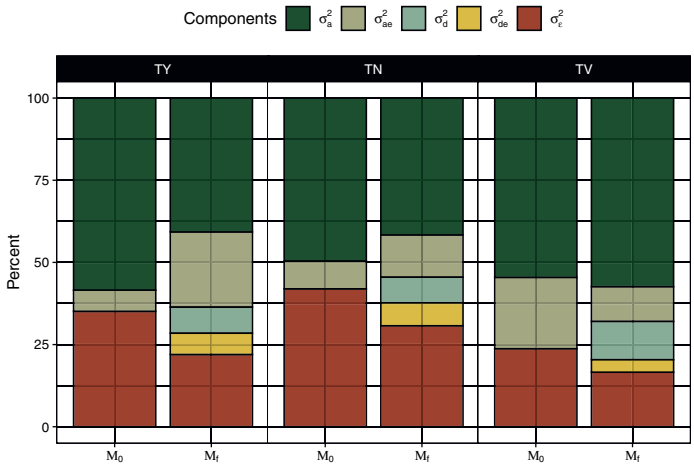


Figure 2.4: Partitioning of total phenotypic variance into additive (σ^2_a), dominant (σ^2_d), interactive (σ^2_{ae} & σ^2_{de}), and residual (σ^2_ε) genetic components across each genetic model (M_0 and M_f) for total yield (TY), tuber number (TN), and average tuber volume (TV).

Table 2.2: The broad and narrow-sense heritabilities (H^2 and h^2, respectively), dominance ratio (d^2), proportion of additive genetic variation ($\frac{\sigma^2_a}{\sigma^2_G}$), and proportion of additive genotype by environment interaction over total genotype by environment interaction ($\frac{\sigma^2_{ae}}{\sigma^2_{GE}}$) estimated from the M_f model for total yield (TY), tuber number (TN), and average tuber volume (TV).

	TY	TN	TV
H^2	0.71	0.74	0.86
h^2	0.59	0.62	0.71
d^2	0.12	0.12	0.14
$\frac{\sigma^2_a}{\sigma^2_G}$	0.84	0.84	0.83
$\frac{\sigma^2_{ae}}{\sigma^2_{GE}}$	0.78	0.65	0.74

Genetic correlations

Along with within-trait variance ratios, genetic correlations were also computed using covariances extracted from model M_f. These were produced for GCA (Figure 2.5) and SCA (Figure 2.6) and are shown together with their BLUP distribution for reference. The GCA intra-class correlation coefficient was found to be highest in tuber volume (0.84) followed by tuber number (0.76) and total yield (0.64) (Figure 2.5). Also noteworthy, large within-trial GCA genetic correlations were found between total yield and each of its yield components, tuber number (0.80) and tuber volume (0.64) according to expectation. Little to no genetic correlation could be found between tuber number and tuber volume (0.06). There were minor discrepancies when comparing each of these with their between-trial genetic correlations counterparts ($\rho_{TY,TN} = 0.81, \rho_{TY,TV} = 0.65, \rho_{TN,TV} = 0.11$).

Similar multivariate trends were observed for SCA genetic correlations, though, with globally smaller values. The SCA intra-class correlation was highest in tuber volume (0.75) with little difference between total yield (0.55) and tuber number (0.53) (Figure 2.6). Relatively large within-trial genetic correlations were observed between total yield and tuber number ($\rho_{TY,TN} = 0.73$) and between total yield and tuber volume ($\rho_{TY,TV} = 0.69$). Genetic correlations between tuber number and tuber volume were virtually null ($\rho_{TN,TV} = 0.03$) showing little to no covariance between these two traits for SCA. Examining between-trait between-trial correlations only had minor deviations with respect to within-trial genetic correlations between total yield and tuber number ($\rho_{TY,TN} = 0.76$) and tuber number and tuber volume ($\rho_{TN,TV} = -0.01$); however, these correlations between total yield and tuber volume were distinctly smaller than their within-trial counter-parts ($\rho_{TY,TV} = 0.55$) (Figure 2.6).

When comparing the GCA and SCA quantiles for each trait and trial location, the GCA BLUPs were consistently larger than those SCA BLUPs. On average, any given GCA quantile was two times larger than its respective SCA quantile; this was true for all traits measured here (See Table S4). These differences in magnitude between the estimated GCA and SCA effects could also be readily seen while examining the size of each variance component (Figure 2.4) or even through a simple regression of hybrid BLUPs on the mid-parent value (Figure 2.7). No linear relationship could be found between the estimated

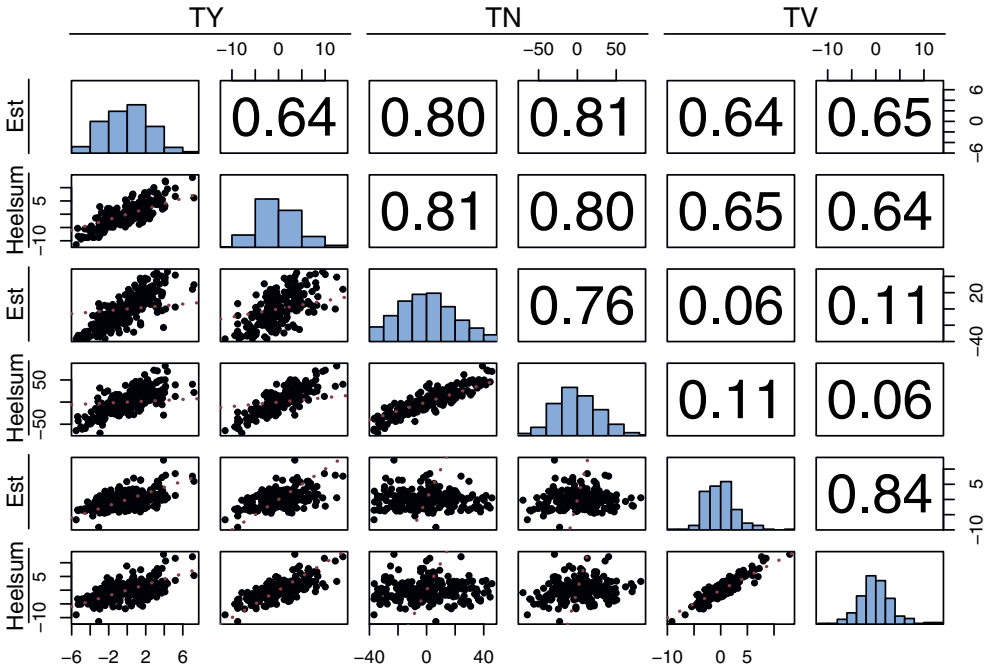


Figure 2.5: General combining ability pairwise comparisons with scatterplot (lower triangle), genetic correlations derived from M_f's variance components (upper triangle), and marginal distributions (the diagonal) of best linear unbiased predictions for total yield (TY; $\frac{Mg}{Ha}$), tuber number (TN; total number of tubers per plot), and average tuber volume (TV; cm^3) in Est and Heelsum. The identity is provided in red for each scatterplot (lower triangle).

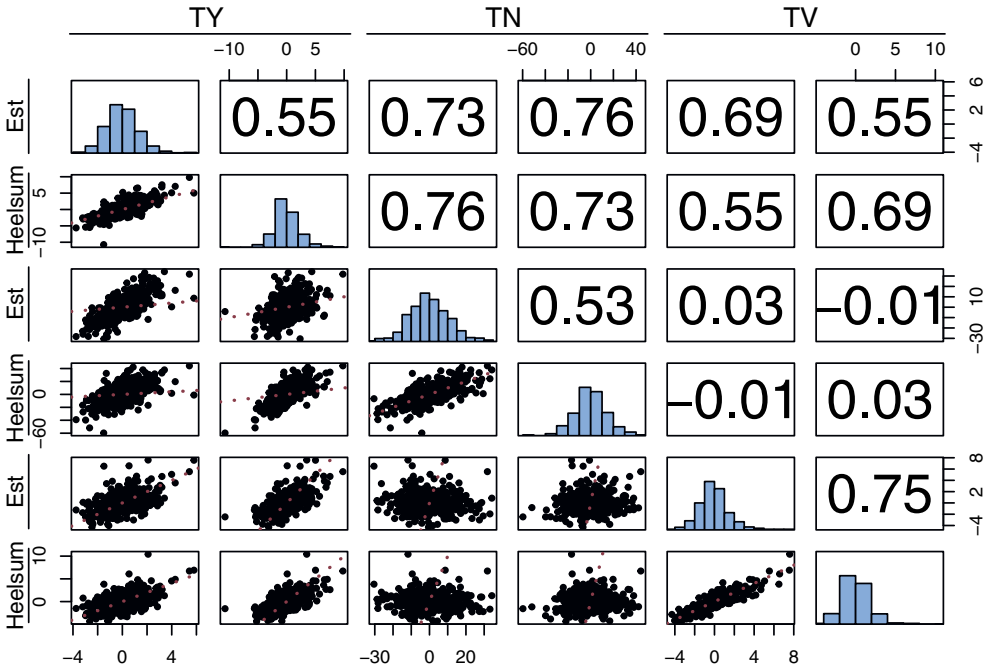


Figure 2.6: Specific combining ability pairwise comparisons with scatterplot (lower triangle), genetic correlations derived from M_f's variance components (upper triangle), and marginal distributions (the diagonal) of best linear unbiased predictions for total yield (TY; $\frac{Mg}{Ha}$), tuber number (TN; total number of tubers per plot), and average tuber volume (TV; cm^3) specific combining ability in Est and Heelsum. The identity is provided in red for each scatterplot (lower triangle).

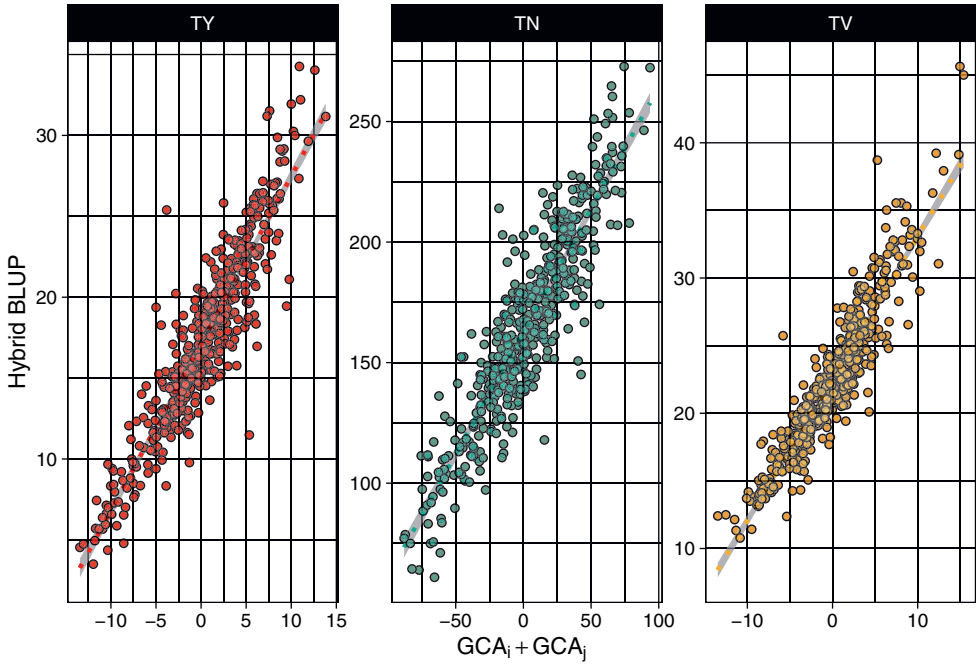


Figure 2.7: Scatterplot of hybrid BLUPs regressed on the GCAs of parents' i and j for total yield (TY; $\frac{Mg}{Ha}$), tuber number (TN; total number of tubers per plot), and average tuber volume (TV; cm^3)

GCA and SCA effects for any trait (See Figure S1) coinciding with our model assumptions in (2.5).

Model Testing

The likelihood ratio tests conducted between model M_f and M_0 was found to be significant yielding a probability less than 0.001 (Table 2.3). This test was also conducted on the univariate equivalent of these models for each trait (M_{df} and M_{d0}) and were all found to be significant as well. Each test conformed with the AICs of each model pair where the smallest AIC was observed in the full genetic model suggesting that the best fit was achieved with the inclusion of SCA and its environmental interaction. This at the very least lend a statistical justification for these non-additive genetic effects in all three tuber traits.

2.5 Discussion

Presence of additive vs. non-additive effects

SCA was detected across all phenotypes (Table 2.3) warranting sufficient evidence that SCA can impact yield, especially in its simpler components, seen primarily in tuber volume (Figure 2.4). However, the magnitude of GCA effects were far greater than the magnitude

Table 2.3: Likelihood ratio tests for M_0 and M_f genetic models together with each model's Akaike's information criterion for the Multi-trait model (MT), total yield (TY), tuber number (TN), and average tuber volume (TV).

	AIC	$\Lambda(y)$	$P(\Lambda(y))^{\dagger}$
MT			
M_0	26665	586	***
M_f	26103		
TY			
M_0	7992	132	***
M_f	7864		
TN			
M_0	15778	92	***
M_f	15690		
TV			
M_0	7008	235	***
M_f	6778		

† ***: $P(\Lambda(y)) < 0.001$

of SCAs estimated across all traits irrespective of the their heritability or size of SCA variance. The GCA quantiles were between 1.4 to 2.5 times larger than their respective SCA quantiles (See Table S4) suggesting a systematic importance of the additive genetic effects in this DHP population. The implications then is that most of the variation in the progeny can be found through the additive genetic variation in the parents. This was most apparent for tuber volume ($h^2 = 0.71$) where regression of hybrid performance on the combination of parental GCAs had the best fit of all traits (Figure 2.7). This illustrates that the parental GCAs (or, identically, the mid-parent value) capture the majority of a hybrid's phenotype.

While this study is the largest of its kind, it is certainly not alone in attempting to decompose genetic components of yield in potato. A number of similar populations were used in both diploid and tetraploid backgrounds with wide-ranging results. Many of these studies came to find little non-additive genetic variation for yield components similar to the results presented here with GCA being the primary component for traits like average tuber weight, tuber shape, total tuber number, and total yield[14,40,41]. These studies utilised either variants of a diallel or factorial crossing schema making the structure of their statistical models not altogether different than the modelling endeavoured here. One exception between models' (2.4) and (2.5) and those used in the aforementioned tetraploid populations is that variance attributed to GCA has a different interpretation due to differences in ploidy and levels of inbreeding. This said, there is a large body of work which also finds SCA to be the largest (and at times, the only) detected effect in several complex traits. The same previously mentioned traits as well as others like incidence of hollow heart and tuber uniformity were also found to be predominately under the control

of SCA[40,42–44]. Most notably,[45] was only able to detect SCA in their partial diallel crosses with no GCA component found for marketable yield and marketable tuber number.

The lack of an empirical consensus on the predominance of GCA and SCA in potato is not very surprising in of itself. The estimation of these parameters are very much contingent on numerous factors including crop ploidy, genetic background, number of parents, degree of environmental stress, and choice of statistical model (to name a few), all of which are prone to change across experiments. Even making comparisons between studies utilising very similar genetic backgrounds can lead to divergent findings[46,47]. While seemingly incoherent, the following does offer some interesting grounds for considering those genetic effects observed here. Many of the aforementioned studies estimated variance components on populations which had undergone significant selection through a recurrent selection schema[44,47] or were themselves the product of strong selection on GCAs in their ancestors[45]. In both cases, less additive genetic variation can be expected among hybrids derived from them leaving non-additive genetic variation to be the predominant genetic effect within their inter-population crosses. Conversely, populations like ours which show ample additive genetic variation might be younger with respect to selection pressure in their ancestors on these traits; though without any formal analysis on population structure this is speculative. Another line of reasoning for the smaller SCAs found here relative to many of the tetraploid studies could be explained by *progressive heterosis* whereby higher order non-additive genetic effects become possible through polyploidy[48]. Recent genomic studies in tetraploid potato support this hypothesis with evidence of a genetic residual effect (which could be explained by tri & quadrigenic dominance) contributing as much as 45% of total genetic variation in potato yield[49]. However, making any meaningful confirmation on the specific role of ploidy in producing non-additive genetic effects is beyond the scope of this present study and only deserves a cursory mention.

Genetic Architecture of yield

Among our two yield components (i.e. tuber number and average tuber volume), we found strong genetic correlations between each and total yield for both the additive (see Figure 2.5) and non-additive genetic effects (see Figure 2.6). Numerous studies have identified these same phenotypes as major determinants of total tuber yield marking them both as strategic heritable targets for breeding[43,50]. Consistent with these studies, tuber number GCAs appeared to impact total yield more than tuber volume with $\rho_{TN,TY}$ equalling 0.80 and a $\rho_{TV,TY}$ of 0.64. SCA genetic correlations behaved similarly with $\rho_{TN,TY}$ equal to 0.73 and $\rho_{TV,TY}$ equal to 0.69. Interestingly, while tuber volume had the greatest stability with respect to both GCA ($ICC_{gca} = 0.84$) and SCA ($ICC_{sca} = 0.75$), the between-trial genetic correlations in $\rho_{TV,TY}$ dropped to 0.55 suggesting less coupling in SCA by trial response with total yield in contrast to the between-trial genetic correlations seen in total yield and tuber number ($\rho_{TN,TY}$=0.76). These additive and non-additive components point to tuber number being the primary determinant of yield in this hybrid population.

Among certain market classes, our two yield components, average tuber volume (or tuber size) and tuber number, often exhibit an inverted relationship due to the physiological and genetic limits of potato. For example,[43] found a genetic correlation of -0.24 among their

panel. Additionally,[51] identified negative cubic trends between tuber number and average tuber weight capturing a majority of variation. Interestingly, no meaningful relationship could be found between these two yield components with respect to additive (Figure 2.5) and non-additive genetic correlations (Figure 2.6). To repeat our previous suspicion, this suggests a lack of directional selection on one of these two traits evident by the lack of genetic constraints between them[52]. Tuber volume had the largest proportion of additive variation (Figure 2.4) and genetic variation in general (Table 2.2) suggesting little to no direct selection on this component of yield. These properties then make this population an interesting candidate for future selection given its genetic potential to be adapted to a variety of tuber types. Another oddity to consider here is that while SCA was detected independently in these two yield components, this did not manifest in the increase of SCA in total yield, but very much the opposite. Others have identified that heterosis in these same yield components were responsible for a geometric increase in gross yield[46]. Further multivariate studies of vigour in diploid potato could further elucidate SCA architecture especially as selection pressure is applied, a key scenario within breeding programmes.

Using GCA & SCA in commercial breeding

The large narrow-sense heritabilities and magnitude of the GCAs found here have major implications for breeders of DHP. To begin, the valid estimation of GCAs further validate the potential of potato in its conversion into a inbred-hybrid crop as purported before[53,54]. Further, the size of the estimated GCAs relative to a hybrid's average performance (Figure 2.7) show that standard breeding designs used in other hybrid crops will likely be just as efficacious in DHP. For example, the use of test crosses, a mainstay in maize breeding, can also be utilised in evaluating performance of potato parental lines for hybrid crosses with reasonable success. These test crosses can be further utilised for model training and be the basis for genomic selection of parents (using breeding values) or hybrid crosses (via mid-parent value); again, a standard-place method in hybrid breeding[55]. To quickly add, while we found little contributions made by SCA, their relevance to breeders do not necessarily end here. Depending on the specific mechanism behind these observed non-additive effects, they could be further exploited in the trait of interest through initial breeding design (e.g. formation of heterotic pools). Heterotic breeding has become a major target for quality trait improvement in other solanaceous crops including chilli pepper[56], eggplant[57], and tomato[58]. This is where potato meets an interesting intersection between the vegetable and agronomic worlds where SCA might play a more valuable role for qualities controlling specific market criteria (e.g. average tuber volume, tuber length & shape) but would be less emphasised in composite and complex traits (e.g. gross yield, starch & protein content) where GCA is the predominant genetic effect at play. Having said this, future work into the biometric mechanism of vigour will be able to lend more wisdom for how breeders should wield this in a hybrid potato breeding programme.

2.6 Conclusion

Limitations

The application of these results should be done with some qualification. One principal limitation of this study can be found in the number of trials conducted. All inference drawn here was based upon only two trials which took place over one season. This limits our findings to one particular year which narrows their interpretive weight and scope. Related to this, this study was performed on a particular composite experimental population and does not necessarily represent the heterotic potential of the tuber-bearing Solanum genepool. Considering all this, this population is an appropriate candidate for the purpose of surveying the presence and potential of heterotic vigour in DHP which was the aim of this current paper.

Future work

Finding evidence for heterotic effects in DHP does not yield much regarding the source of the effects identified here. The statistical models assume all underlying effects captured by the SCA term to be the product of cumulative dominance deviations across the genome, but there exists many other plausible sources of non-additive variation. Since its conception, genetic theory has explained heterosis with a whole suite of models with many being broadly plausible (See[59]). However, these apparent non-additive effects could just as simply be explained by dispersion of additive alleles among parents, a hypothesis which is generally supported empirically[58,60]. Nevertheless, these effects continue to be interesting point of study and still remain an important target in hybrid breeding of modern crops. This is especially worthwhile in DHP given its novelty as a hybrid crop with the potential of heterotic breeding to still be determined.

This is the first study in diploid hybrid potato to produce estimates of general and specific combining abilities using a large panel of commercially-derived parents. This represents a major milestone in the reorientation of potato from a clonal tetraploid to a diploid inbred-hybrid crop. Identifying the predominance of additive genetic effects for multiple yield components among hybrids offers strategic insight on the necessity of effective generation of parental lines and early population development in general. Though the estimated non-additive effects in this population are smaller in contrast to their additive counterparts, heterotic vigour shows some minor role in simpler traits. Specific quantitative traits should be targeted for SCA exploitation to bolster variety development on top of their parental effects. Further research into the genetic mechanisms for the apparent non-additive effects will also better elucidate the strategic advantage (if any exist) in key economic targets in hybrid potato.

Data availability

All trial and pedigree data utilized for the following analysis have been made available on GSA figshare (https://doi.org/10.25387/g3.16973293). File Phenotypes.csv contains

the 3 aforementioned traits along with field trial row, column, and block indices for each observation. File Pedigrees.csv gives a hybrid identification number with each parental code.

Funding

The research produced for this publication was made possible through funding provided by Topsector Tuinbouw & Uitgangsmaterialen in the 'Re-booting potato: enhancing the breeding of hybrid diploid potato' project (project number: TU-18155).

Bibliography

1. Adams JR, de Vries ME, Zheng C, and van Eeuwijk FA. Little Heterosis Found in Diploid Hybrid Potato: The Genetic Underpinnings of a New Hybrid Crop. G3: Genes, Genomes, Genetics. 2022 Jun; 12. DOI: 10.1093/g3journal/jkac076. PMID: 35460241
2. FAOSTAT. Food and Agriculture Organization of the United Natons. FAOSTAT database, 2021. Available from: http://www.fao.org/faostat
3. Hay RK. Harvest Index: A Review of Its Use in Plant Breeding and Crop Physiology. Annals of Applied Biology. 1995; 126:197–216. DOI: 10.1111/j.1744-7348.1995.tb05015.x
4. Lutaladio NB and Castaldi L. Potato: The Hidden Treasure. Journal of Food Composition and Analysis. 2009 Sep 1; 22:491–3. DOI: 10.1016/j.jfca.2009.05.002
5. Douches DS, Maas D, Science KJC, and 1996 U. Assessment of Potato Breeding Progress in the USA over the Last Century. Crop Science. 1996; 36:1544–52. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/36/6/CS0360061544 [Accessed on: 2019 Apr 29]
6. Hirsch CN, Hirsch CD, Felcher K, Coombs J, Zarka D, Van Deynze A, De Jong W, Veilleux RE, Jansky S, Bethke P, Douches DS, and Buell CR. Retrospective View of North American Potato (Solanum Tuberosum L.) Breeding in the 20th and 21st Centuries. G3: Genes, Genomes, Genetics. 2013; 3:1003–13. DOI: 10.1534/g3.113.005595
7. Lian Q, Tang D, Bai Z, Qi J, Lu F, Huang S, and Zhang C. Acquisition of Deleterious Mutations during Potato Polyploidization. Journal of Integrative Plant Biology. 2019 Jan 1; 61:7–11. DOI: 10.1111/jipb.12748
8. Zhang C, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, and Huang S. The Genetic Basis of Inbreeding Depression in Potato. Nature Genetics. 2019 Mar 14; 51:374–8. DOI: 10.1038/s41588-018-0319-1. Available from: http://www.nature.com/articles/s41588-018-0319-1 [Accessed on: 2019 Mar 20]
9. Bartlett MS and Haldane JB. The Theory of Inbreeding in Autotetraploids. Journal of Genetics. 1934 Jul; 29:175–80. DOI: 10.1007/BF02982194
10. Simmonds N. Principles of Crop Improvement. New York: Longman Group, 1979. 123-204
11. Bradshaw JE. Review and Analysis of Limitations in Ways to Improve Conventional Potato Breeding. Potato Research. 2017; 60:171–93. DOI: 10.1007/s11540-017-

9346-z. Available from: http://www.tsl.ac.uk/news/new-potato-at-the- [Accessed on: 2021 Jun 28]
12. Bryan JE. Rapid Multiplication Techniques for Potatoes. Lima, 1981 :3–20. Available from: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=LIBROS.xis&method=post&formato=2&cantidad=1&expresion=mfn=008106 [Accessed on: 2021 Jun 28]
13. Tai GC and Young DA. Early Generation Selection for Important Agronomic Characteristics in a Potato Breeding Population. American Potato Journal. 1984 Jul; 61:419–34. DOI: 10.1007/BF02852811
14. Neele AEF, Nab HJ, and Louwes KM. Identification of Superior Parents in a Potato Breeding Programme. Theoretical and Applied Genetics. 1991 Sep; 82:264–72. DOI: 10.1007/BF02190611. Available from: https://link.springer.com/article/10.1007/BF02190611 [Accessed on: 2021 Oct 9]
15. Bradshaw JE, Dale ·MFB, and Mackay ·GR. Use of Mid-Parent Values and Progeny Tests to Increase the Efficiency of Potato Breeding for Combined Processing Quality and Disease and Pest Resistance. Theor Appl Genet. 2003; 107:36–42. DOI: 10.1007/s00122-003-1219-y
16. Hougas R and Peloquin S. The Potential of Potato Haploids in Breeding and Genetic Research. American Potato Journal. 1958; 35:701–7
17. none and Cock JH. Global Workshop on Root and Tuber Crops Propagation. Cali, Colombia, 1983
18. Pallais N. True Potato Seed: Changing Potato Propagation from Vegetative to Sexual. HortScience. 1991 Mar; 26:239–41
19. Thomas-Sharma S, Abdurahman A, Ali S, Andrade-Piedra JL, Bao S, Charkowski AO, Crook D, Kadian M, Kromann P, Struik PC, Torrance L, Garrett KA, and Forbes GA. Seed Degeneration in Potato: The Need for an Integrated Seed Health Strategy to Mitigate the Problem in Developing Countries. Plant Pathology. 2016 Jan 1; 65:3–16. DOI: 10.1111/ppa.12439. Available from: https://bsppjournals.onlinelibrary.wiley.com/doi/full/10.1111/ppa.12439 [Accessed on: 2021 Jun 25]
20. Hosaka K and Hanneman RE. Genetics of Self-Compatibility in a Self-Incompatible Wild Diploid Potato Species Solanum Chacoense. 2. Localization of an S Locus Inhibitor (Sli) Gene on the Potato Genome Using DNA Markers *. Euphytica. 1998; 103:265–71
21. Alsahlany M, Enciso-Rodriguez F, Lopez-Cruz M, Coombs J, and Douches DS. Developing Self-Compatible Diploid Potato Germplasm through Recurrent Selection. Euphytica. 2021 Mar 1; 217:1–16. DOI: 10.1007/s10681-021-02785-0. Available from: https://doi.org/10.1007/s10681-021-02785-0 [Accessed on: 2021 Jun 28]
22. Stockem J, de Vries M, van Nieuwenhuizen E, Lindhout P, and Struik PC. Contribution and Stability of Yield Components of Diploid Hybrid Potato. Potato Research. 2020 Sep 1; 63:345–66. DOI: 10.1007/s11540-019-09444-x. Available from: https://doi.org/10.1007/s11540-019-09444-x [Accessed on: 2021 Feb 2]

23. Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, and Huang S. Genome Design of Hybrid Potato. Cell. 2021 Jul 22; 184:3873–3883.e12. DOI: 10.1016/J.CELL.2021.06.006
24. Lindhout P, de Vries M, ter Maat M, Ying S, Viquez-Zamora Marcela, and van Deusden S. Hybrid Potato Breeding for Improved Varieties. *Achieving Sustainable Cultivation of Potatoes Volume 1*. Ed. by Wang-Pruski Gefu. Burleigh Dodds Science Publishing, 2018 Apr 9:119–42. DOI: 10.1201/9781351114455-10/HYBRID-POTATO-BREEDING-IMPROVED-VARIETIES-GEFU-WANG-PRUSKI
25. Kacheyo OC, van Dijk LCM, de Vries ME, and Struik PC. Augmented Descriptions of Growth and Development Stages of Potato (Solanum Tuberosum L.) Grown from Different Types of Planting Material. Annals of Applied Biology. 2021 May 1; 178:549–66. DOI: 10.1111/aab.12661. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/aab.12661 [Accessed on: 2021 Jun 8]
26. Stockem JE, Korontzis G, Wilson SE, Vries ME de, Eeuwijk FA, and Struik PC. Optimal Plot Dimensions for Performance Testingfof Hybrid Potato in the Field. Potato Research. 2021
27. Rodriguez-Alvarez MX, Boer MP, van Eeuwijk FA, and Eilers PH. Correcting for Spatial Heterogeneity in Plant Breeding Experiments with P-splines. Spatial Statistics. 2018; 23:52–71. Available from: https://www.sciencedirect.com/science/article/pii/S2211675317301070 [Accessed on: 2025 May 17]
28. Velazco JG, Rodríguez-Álvarez MX, Boer MP, Jordan DR, Eilers PH, Malosetti M, and van Eeuwijk FA. Modelling Spatial Trends in Sorghum Breeding Field Trials Using a Two-Dimensional P-spline Mixed Model. Theoretical and Applied Genetics. 2017 Jul 1; 130:1375–92. DOI: 10.1007/s00122-017-2894-4. PMID: 28374049. Available from: https://cran.r-project.org/ [Accessed on: 2021 Jun 8]
29. Lee DJ, Durbán M, and Eilers P. Efficient Two-Dimensional Smoothing with P - Spline ANOVA Mixed Models and Nested Bases. Computational Statistics & Data Analysis. 2013 May; 61:22–37. DOI: 10.1016/j.csda.2012.11.013. Available from: https://linkinghub.elsevier.com/retrieve/pii/S016794731200415X [Accessed on: 2025 May 17]
30. Oakey H, Verbyla A, Pitchford W, Cullis B, and Kuchel H. Joint Modeling of Additive and Non-Additive Genetic Line Effects in Single Field Trials. Theoretical and Applied Genetics. 2006 Sep; 113:809–19. DOI: 10.1007/s00122-006-0333-z. PMID: 16896718. Available from: https://pubmed.ncbi.nlm.nih.gov/16896718/ [Accessed on: 2021 Jan 20]
31. Sprague GF and Tatum LA. General vs. Specific Combining Ability in Single Crosses of Corn. Agronomy Journal. 1942 Oct; 34:923–32. DOI: 10.2134/AGRONJ1942.00021962003400100008X
32. Bernardo R. Breeding for Quantitative Traits in Plants. Woodbury, Stemma Press, 2002. Available from: http://stemmapress.com [Accessed on: 2021 Jun 29]
33. Eisenhart C. The Assumptions Underlying the Analysis of Variance. Biometrics. 1947 Mar; 3:1–21. DOI: 10.2307/3001534. PMID: 20240414
34. Griffing B. Concept of General and Specific Combining Ability in Relation to Diallel Crossing Systems. Australian journal of biological sciences. 1956; 9:463–93

35. Xenakis JG. Special Topics For Recombinant Inbred Intercross Data: Model Identifiability, Hypothesis Testing and Compositional Methods. The University of North Carolina at Chapel Hill, 2019. Available from: https://search.proquest.com/openview/28f4f585a836499e5a03e93f37b54d04/1?pq-origsite=gscholar&cbl=18750&diss=y [Accessed on: 2021 Jul 6]
36. Self SG and Liang KY. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. Journal of the American Statistical Association. 1987; 82:605–10. DOI: 10.1080/01621459.1987.10478472
37. Butler DG, Cullis BR, Gilmour AR, Gogel BJ, and Thompson R. ASReml-R Reference Manual Version 4. Hemel Hempstead, HP1 1ES: VSN International Ltd, 2017. Available from: http://www.homepages.ed.ac.uk/iwhite/asreml/uop. [Accessed on: 2021 Jan 19]
38. Searle S, Casella G, and McCulloch C. Variance Components. Ed. by Searle SR, Casella G, and McCulloch CE. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1992 Mar 13. 249-255. DOI: 10.1002/9780470316856. Available from: http://doi.wiley.com/10.1002/9780470316856 [Accessed on: 2021 Feb 2]
39. Möhring J, Melchinger AE, and Piepho HP. REML-Based Diallel Analysis. Crop Science. 2011 Mar 1; 51:470–8. DOI: 10.2135/cropsci2010.05.0272. Available from: http://doi.wiley.com/10.2135/cropsci2010.05.0272 [Accessed on: 2021 Feb 3]
40. Veilleux RE and Lauer FI. Breeding Behavior of Yield Components and Hollow Heart in Tetraploid-Diploid vs. Conventionally Derived Potato Hybrids. Euphytica 1981 30:3. 1981; 30:547–61. DOI: 10.1007/BF00038781. Available from: https://link.springer.com/article/10.1007/BF00038781 [Accessed on: 2021 Oct 9]
41. Brown J and Caligari PDS. Cross Prediction in a Potato Breeding Programme by Evaluation of Parental Material. Theoretical and Applied Genetics. 1989 Feb; 77:246–52. DOI: 10.1007/BF00266194. Available from: https://link.springer.com/article/10.1007/BF00266194 [Accessed on: 2021 Oct 9]
42. Killick RJ. Genetic Analysis of Several Traits in Potatoes by Means of a Diallel Cross. Annals of Applied Biology. 1977 Jul 1; 86:279–89. DOI: 10.1111/J.1744-7348.1977.TB01841.X. Available from: https://onlinelibrary-wiley-com.ezproxy.library.wur.nl/doi/full/10.1111/j.1744-7348.1977.tb01841.x [Accessed on: 2021 Oct 9]
43. Thompson PG and Mendoza HA. Genetic Variance Estimates in a Heterogenous Potato Population Propagated from True Seed (TPS). American Potato Journal. 1984 Nov; 61:697–702. DOI: 10.1007/BF02852933. Available from: https://link.springer.com/article/10.1007/BF02852933 [Accessed on: 2021 Mar 6]
44. Haynes KG. Variance Components for Yield and Specific Gravity in a Diploid Potato Population after Two Cycles of Recurrent Selection. 2001 :69–75
45. Tai G. ESTIMATION OF GENERAL AND SPECIFIC COMBINING ABILITIES IN POTATO. Canadian Journal of Genetics and Cytology. 1976 Sep 1; 18:463–70. DOI: 10.1139/g76-056. Available from: http://www.nrcresearchpress.com/doi/10.1139/g76-056 [Accessed on: 2025 May 17]

46. Tarn TR and Tai GCC. Heterosis and Variation of Yield Components in F1 Hybrids Between Group Tuberosum and Group Andigena Potatoes. Crop Science. 1977; 17:cropsci1977.0011183X001700040009x. DOI: 10.2135/cropsci1977.0011183X001700040009x. Available from: https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci1977.0011183X001700040009x [Accessed on: 2025 May 17]
47. Maris M. Analysis of an Incomplete Diallel Cross among Three Ssp. Tuberosum Varieties and Seven Long-Day Adapted Ssp. Andigena Clones of the Potato (Solanum Tuberosum L.) Euphytica. 1989 Apr; 41:163–82. DOI: 10.1007/BF00022425. Available from: https://link.springer.com/article/10.1007/BF00022425 [Accessed on: 2021 Oct 9]
48. Birchler JA, Auger DL, and Riddle NC. In Search of the Molecular Basis of Heterosis. The Plant Cell. 2003; 15:2236–9. Available from: https://academic.oup.com/plcell/article-abstract/15/10/2236/6010127 [Accessed on: 2025 May 17]
49. Endelman JB, Carley CA, Bethke PC, Coombs JJ, Clough ME, da Silva WL, De Jong WS, Douches DS, Frederick CM, Haynes KG, Holm DG, Miller JC, Muñoz PR, Navarro FM, Novy RG, Palta JP, Porter GA, Rak KT, Sathuvalli VR, Thompson AL, and Yencho GC. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics. 2018 May 1; 209:77–87. DOI: 10.1534/genetics.118.300685. PMID: 29514860. Available from: https://pubmed.ncbi.nlm.nih.gov/29514860/ [Accessed on: 2021 Jan 29]
50. Khayatnezhad M, Shahriari R, Gholamin R, Jamaati-E-Somarin S, and Zabihi-E-Mahmoodabad R. Correlation and Path Analysis Between Yield and Yield Components in Potato (Solanum Tubersum L.) Middle-East Journal of Scientific Research. 2011; 7:17–21. Available from: https://www.researchgate.net/publication/338051404 [Accessed on: 2021 Aug 12]
51. Lemaga B and Caesar K. Relationships between Numbers of Main Stems and Yield Components of Potato (Solanum Tuberosum L. Cv. Erntestolz) as Influenced by Different Daylengths. Potato Research. 1990 Jun; 33:257–67. DOI: 10.1007/BF02358455. Available from: https://link.springer.com/article/10.1007/BF02358455 [Accessed on: 2021 Jun 28]
52. Blows M and Walsh B. Spherical Cows Grazing in Flatland: Constraints to Selection and Adaptation. *Adaptation and Fitness in Animal Populations*. Ed. by van der Werf J, Graser HU, Frankham R, and Gondro C. Dordrecht: Springer, 2009 :83–101. DOI: 10.1007/978-1-4020-9005-9_6
53. Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS, Bamberg JB, Thompson AL, Bizimungu B, Holm DG, Brown CR, Haynes KG, Sathuvalli VR, Veilleux RE, Creighton Miller J, Bradeen JM, and Jiang J. Reinventing Potato as a Diploid Inbred Line-Based Crop. Crop Science. 2016 Jul 1; 56:1412–22. DOI: 10.2135/cropsci2015.12.0740
54. Lindhout P, Meijer D, Schotte T, Hutten RC, Visser RG, and Eck HJ van. Towards F1 hybrid seed potato breeding. Potato Research 2011; 54:301–12

55. Albrecht T, Wimmer V, Rgen Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, and Schön CC. Genome-Based Prediction of Testcross Values in Maize. Theoretical and Applied Genetics. 2011; 123:339–50. DOI: 10.1007/s00122-011-1587-7
56. Herath HMSN, Rafii MY, Ismail SI, JJ. N, and Ramlee SI. Improvement of Important Economic Traits in Chilli through Heterosis Breeding: A Review. https://doiorg/101080/1462031620201780162. 2020; 96:14–23. DOI: 10.1080/14620316.2020.1780162. Available from: https://www.tandfonline.com/doi/abs/10.1080/14620316.2020.1780162 [Accessed on: 2021 Aug 11]
57. Kumar A, Sharma V, Jain BT, and Kaushik P. Heterosis Breeding in Eggplant (Solanum Melongena L.): Gains and Provocations. Plants 2020, Vol 9, Page 403. 2020 Mar 24; 9:403. DOI: 10.3390/PLANTS9030403. Available from: https://www.mdpi.com/2223-7747/9/3/403/htm [Accessed on: 2021 Aug 11]
58. none. Heterosis Reappraisal of Theory and Practice. Ed. by Frankel R. Springer, Berlin, Heidelberg, 1983. DOI: 10.1007/978-3-642-81977-3_6. Available from: https://link.springer.com/chapter/10.1007/978-3-642-81977-3_6 [Accessed on: 2021 Jun 24]
59. Labroo MR, Studer AJ, and Rutkoski JE. Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. Frontiers in Genetics. 2021 Feb 24; 12:234. DOI: 10.3389/FGENE.2021.643761
60. Mackay IJ, Cockram J, Howell P, and Powell W. Understanding the Classics: The Unifying Concepts of Transgressive Segregation, Inbreeding Depression and Heterosis and Their Central Relevance for Crop Breeding. Plant Biotechnology Journal. 2021 Jan 1; 19:26–34. DOI: 10.1111/PBI.13481. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/pbi.13481 [Accessed on: 2021 Oct 14]

2.7 Supplementary

Table S1: Spatial model summaries for Est describing effective and nominal dimensions, their ratio, and parameter type across all traits

	TY				TN				TV			
	Effective Dimension	Nominal Dimension	Ratio	Type[a]	Effective Dimension	Nominal Dimension	Ratio	Type[a]	Effective Dimension	Nominal Dimension	Ratio	Type[a]
Intercept	1.0	1	1.00	F	1.0	1	1.00	F	1.0	1	1.00	F
Control	1.0	1	1.00	F	1.0	1	1.00	F	1.0	1	1.00	F
Genotype	505.5	604	0.84	R	483.0	604	0.80	R	539.6	604	0.89	R
R-Row	12.2	18	0.68	R	12.2	18	0.68	R	8.5	18	0.47	R
R-Column	0.0	63	0.00	R	0.0	63	0.00	R	0.0	63	0.00	R
Column	1.0	1	1.00	S	1.0	1	1.00	S	1.0	1	1.00	S
Row	1.0	1	1.00	S	1.0	1	1.00	S	1.0	1	1.00	S
Column:Row	1.0	1	1.00	S	1.0	1	1.00	S	1.0	1	1.00	S
f(Column)	2.9	21	0.14	S	3.4	21	0.16	S	0.9	21	0.04	S
f(Row)	0.0	66	0.00	S	0.0	66	0.00	S	0.0	66	0.00	S
f(Column):Row	3.8	21	0.18	S	0.0	21	0.00	S	3.0	21	0.14	S
Column:f(Row)	0.7	66	0.01	S	0.6	66	0.01	S	0.0	66	0.00	S
f(Column):f(Row)	16.5	1386	0.01	S	11.7	1386	0.01	S	2.9	1386	0.00	S
Total	546.7	2250	0.24		515.9	2250	0.23		559.8	2250	0.25	
Residual	635.3				666.1				622.2			

[a] F = Fixed, R = Random, S = Semi-parametric

Table S2: Spatial model summaries for Heelsum describing effective and nominal dimensions, their ratio, and parameter type across all traits

	TY				TN				TV			
	Effective Dimension	Nominal Dimension	Ratio	Type[a]	Effective Dimension	Nominal Dimension	Ratio	Type[a]	Effective Dimension	Nominal Dimension	Ratio	Type[a]
Intercept	1.0	1	1.00	F	1.0	1	1.00	F	1.0	1	1.00	F
Control	1.0	1	1.00	F	1.0	1	1.00	F	1.0	1	1.00	F
Genotype	634.0	718	0.88	R	585.6	718	0.82	R	643.0	718	0.90	R
R-Row	0.0	18	0.00	R	0.0	18	0.00	R	0.0	18	0.00	R
R-Column	37.3	83	0.45	R	39.7	83	0.48	R	45.3	83	0.55	R
Column	1.0	1	1.00	S	1.0	1	1.00	S	1.0	1	1.00	S
Row	1.0	1	1.00	S	1.0	1	1.00	S	1.0	1	1.00	S
Column:Row	1.0	1	1.00	S	1.0	1	1.00	S	1.0	1	1.00	S
f(Column)	4.5	21	0.21	S	3.2	21	0.15	S	2.2	21	0.10	S
f(Row)	0.0	86	0.00	S	0.4	86	0.01	S	0.9	86	0.01	S
f(Column):Row	0.0	21	0.00	S	0.2	21	0.01	S	0.0	21	0.00	S
Column:f(Row)	0.0	86	0.00	S	0.3	86	0.00	S	0.0	86	0.00	S
f(Column):f(Row)	18.3	1806	0.01	S	13.3	1806	0.01	S	6.4	1806	0.00	S
Total	699.2	2844	0.25		647.6	2844	0.23		702.8	2844	0.25	
Residual	771.8				823.4				731.2			

[a] F = Fixed, R = Random, S = Semi-parametric

Table S3: Variance components from hybrid model with standard errors included in parentheticals for total tuber number (TN), average tuber volume (TV), and total tuber yield (TY). Covariances between traits excluded here for brevity.

Component	TN	TV	TY
σ^2_G	1432.98 (131.33)	27.29 (2.04)	25.60 (2.42)
σ^2_{GE}	589.55 (73.30)	5.27 (0.60)	15.06 (1.41)
σ^2_ε	923.07 (43.72)	6.66 (0.32)	11.84 (0.56)

Table S4: Ratio between the sample quantiles of GCA and SCA effects for total yield (TY), tuber number (TN), and average tuber volume (TV). Included are the 0, 10th, 25th, 40th, 60th, 75th, 90th and 100th percentiles along with an average ratio of quantiles in $\overline{q}$.

Trait	0%	10%	25%	40%	60%	75%	90%	100%	$\overline{q}$
TN	1.4	2.0	2.4	2.4	2.5	2.1	2.2	1.6	2.1
TV	2.5	2.0	1.7	1.6	1.8	2.1	1.7	1.4	1.9
TY	1.4	2.4	2.0	2.4	2.3	2.5	2.0	1.4	2.0

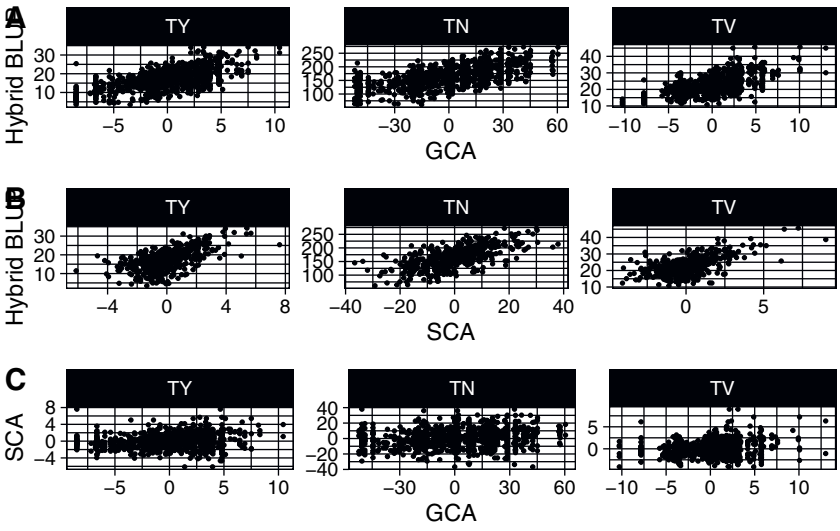


Figure S1: Scatterplots for each tuber trait with (A) hybrid BLUPs plotted with GCAs, (B) SCAs, and (C) GCAs on SCAs

3

Efficient genomic prediction of yield and dry matter in hybrid potato

This chapter is based on:

Adams J, De Vries M, and Van Eeuwijk F. Efficient Genomic Prediction of Yield and Dry Matter in Hybrid Potato. Plants. 2023 Jul 11; 12:2617. doi: 10.3390/plants12142617. Available from: https://www.mdpi.com/2223-7747/12/14/2617 [Accessed on: 2023 Nov 9]

Abstract

There is an ongoing endeavor within the potato breeding sector to rapidly adapt potato from a clonal polyploid crop to a diploid hybrid potato crop. While hybrid breeding allows for efficient generation and selection of parental lines, it also increases breeding program complexity and results in longer breeding cycles. Over the past two decades, genomic prediction has revolutionized hybrid crop breeding through shorter breeding cycles, lower phenotyping costs, and better population improvement resulting in increased genetic gains for genetically complex traits. In order to accelerate genetic gains in hybrid potato, proper implementation of genomic prediction is a crucial milestone in the rapid improvement for this crop. The authors of this paper set out to test genomic prediction in hybrid potato using current genotyped material with two alternative models: one model that predicts the general combining ability effects (GCA) and another which predicts both general and specific combining ability effects (GCA+SCA). Using a training set comprised of 769 hybrids and 456 genotyped parental lines, we found that reasonable prediction accuracies could be achieved for most phenotypes with both zero common parents ($\rho = 0.36 - 0.61$) and one ($\rho = 0.50 - 0.68$) common parent between training and test sets. There was no benefit with the inclusion of non-additive genetic effects in the GCA+SCA model despite SCA variance contributing between 9% and 19% of total genetic variance. Genotype by environment interaction, while present, did not appear to affect prediction accuracies though prediction errors did vary across trial targets. These results suggest that genomically estimated breeding values on parental lines are sufficient for hybrid yield prediction.

Hybrid potato; genomic prediction; general combining ability; genomically estimated breeding values; Hybrid prediction; genomic prediction; Genotype by environment

3.1 Introduction

Potato (*Solanum tuberosum* L.) is the most important starch source among vegetable crops. Cultivated for its edible tubers, potato's end-use spans across multiple markets including table, chipping, frying, and industrial processing with a global export value of 3.36 billion USD[2]. It has positive production characteristics including excellent water use efficiency and short crop cycles making it amenable to a variety of cropping systems[3]. Despite potato's potential for food systems worldwide, slower genetic progress, particularly for quantitative traits such as yield, is a known impediment to long-term food security[4]. Underlying factors are numerous, but this is understood to be the consequence, in part, by potato's polyploidy and large genetic load[5]. In the context of breeding, this makes potato an obligate outcrosser and affects the whole breeding system from initial crosses to the selection schemas used in identifying favourable varietal candidates. Together with potato's large selection surface (as many as 40 traits), improving quantitative traits in clonal potato is rife with technical difficulties[6,7].

There have been many proposed (and implemented) innovations for conventional clonal breeding including the use of marker-assisted selection, progeny testing strategies, and genetic transformation[8]. Among these, one topic which has garnered significant interest over the past few decades is adapting tetraploid potato ($2n = 4x = 48$) to a diploid hybrid breeding system ($2n = 2x = 24$)[9,10]. The appeal of diploid hybrid potato (DHP) breeding is understood to consist of benefits such as the use of true potato seed (TPS) directly in varietal evaluation[11], segmented breeding program design[12], exploitation of heterotic vigour[13], and more efficient systems for germplasm production and dissemination[14]. With the gradual emergence of fertile diploid lines[15,16], research centered around diploid and hybrid potato breeding are gaining momentum across the world[10,17–19].

With all of hybrid breeding's benefits, it is also known for being more resource and time intensive in contrast to inbred systems[20]. Hybrid breeding systems are also more complex in terms of breeding targets; not only must the breeder find the best set of parents (targeting general combining ability, i.e., GCA), but also find the best cross among that set of parents (targeting specific combining ability, i.e., SCA). Because of this, genomic prediction (GP) and genomic selection (GS) have become indispensable methods in the implementation of hybrid breeding in multiple crops[20,21]. The wide-spread adoption of these techniques have been made possible through the advent of high-throughput molecular marker data and the availability of powerful computational resources capable of rapid model fitting[22,23]. These methods are conceptual extensions of marker-assisted selection where rather than selecting an individual on its QTL status, selection is on the basis of genome-wide estimated breeding values (GEBV's). GP and GS have been used to increase the efficiency of hybrid breeding through better heterotic pool development[24], intra-population improvement[25], and even shortening breeding cycles by forgoing trialling altogether[26]. Because of GP and GS's demonstrated gains in hybrid, and crop breeding

more generally, developing appropriate models for each target trait with adequate training sets is a major pursuit in any breeding endeavour.

Recently, GP has received much attention in tetraploid potato with encouraging results demonstrated in chipping quality[27,28], disease resistance[29,30],fry colour[31], dry matter content and specific gravity[27,32], yield components[30,32–34], and vine maturity[28,35]. Despite the increased interest in breeding DHP, there are currently no analogous studies confirming model performance for economically valuable traits. Additionally, while targeting the general combining ability (GCA) gives acceptable prediction accuracies in both simulation studies and *in situ* for many crops, evaluating the contribution of non-additive effects (often through specific combining ability, i.e., SCA) is an altogether crucial question in GP's implementation. The authors of this study wish to showcase GP in DHP with three primary aims. First, to demonstrate the feasibility of GP for hybrid performance of multiple yield components and tuber dry matter content. The second, to test for the contribution of additive and non-additive genetic effects for each tuber phenotype. Lastly, a cursory examination on genotype by environment effects to understand their impact on prediction among the trials used in this study. Prediction models were constructed by adapting the modelling procedure of[36] by structuring the random genetic effects with marker information. Through this work, we touch upon several facets of predictive breeding, all of which, are crucial for adoption and improvement of hybrid potato within the wider potato industry.

3.2 Material and methods

Populations & Trials

Between 2017 and 2020, several crossing blocks were constructed around 712 experimental inbred lines sourced from Solynta, a Dutch hybrid potato breeder[17]. These lines were generated from an original population comprised of 16 founders and selected primarily based on tuber criteria as well as on fertility characteristics during inbreeding. Across five nurseries, 1,162 hybrids were produced through a sparse mating design, i.e., only a small fraction of crosses were conducted over the full crossing space (n^2)[37]. Hybrid TPS from these crosses were sown in trays in early-March and transplanted in mid-May in both trial years. These hybrids were transplanted in all four field trials in two trial locations (Heelsum and Est, Netherlands) in 2019 and 2020. The trials conducted in 2019 tested the performance of 806 F1 crosses, as previously described in[36], with 356 additional hybrids evaluated in 2020 with a thin intersect of hybrids between years (43 hybrids) and thinner intersect across all trials (25 hybrids). All trials utilised a double ridge design comprising of 16 plants per plot; ridges were 75 cm wide with plants spaced 25 cm within the same ridge. Each trial utilised a randomised complete block design consisting of two replicates for each hybrid with additional control varieties planted in diagonals across blocks for spatial trend control. Haulm killing took place in September with subsequent harvests occurring two weeks later. All tubers excluding those smaller than the 20 mm size class (due to instrumentation constraints) were harvested and measured on multiple traits including total tuber yield (TY; $\mathrm{M}g \cdot \mathrm{Ha}^{-1}$), total tuber number (TN; number of tubers per plot)),

percentage of tuber dry matter content (DM; per cent), and average tuber volume (TV; cm^3). All phenotyping procedures, trait definitions, and instrumentation used here are described in[38].

Marker Processing and Analysis

The parental lines used for test-crosses were genotyped using a targeted genotyping-by-sequencing (GBS) technology, SeqSNP®[39]. These probes were designed around highly conserved regions across the genome. Per LGC's protocol, leaf disks were sampled from each inbred line and placed in a 96-well plate and sent to LGC for DNA extraction and analysis. Upon reception of results, marker quality was assessed primarily on confirming coverage and read depth for each sample. For the purpose of this study, these probes were filtered for high-quality biallelic SNP's using multiple criteria. These included typical filtering criteria such as minor allele frequency ($p \geq 0.05$), percentage of SNP missingness ($m \leq 0.1$) along with polymorphism information content (PIC) and linkage disequilibrium (LD) between sites. The latter was chosen due to high LD between many SNP's resulting in redundant marker information. A SNP selection procedure was invoked using the r_v^2 metric provided by[40] which accounts for shared population history which tends to bias the traditional r^2 statistic. For any site in high LD with another ($r_v^2 \geq 0.5$), the site with the highest PIC was chosen. This resulted in a filtered probe set containing 704 SNP's. There was a fraction of parental lines that were either genotyped with an irrelevant probe set or not genotyped at all. This resulted in 456 of the original 712 parental lines being present in the marker set.

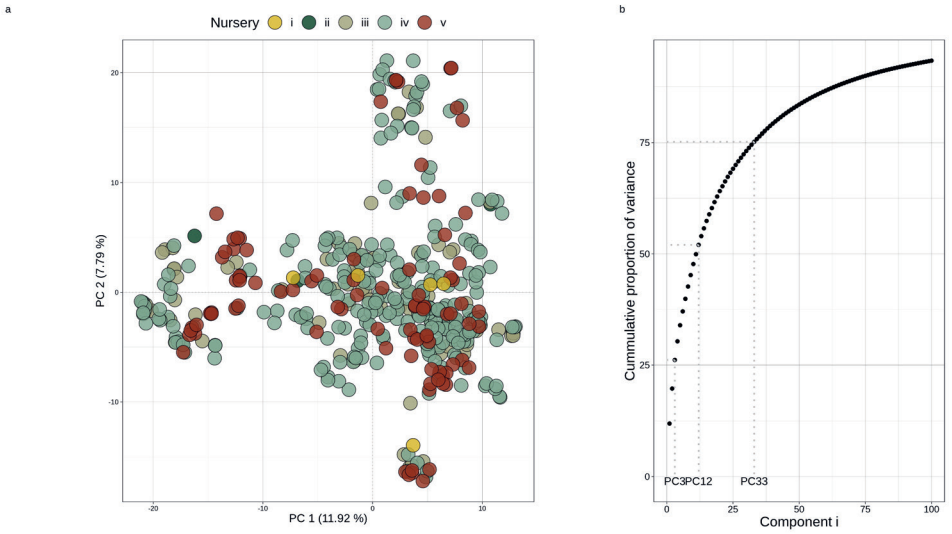


Figure 3.1: (a) Principal components 1 and 2 using filtered marker data (704 SNPs) for 456 parental lines. Parental lines are coloured per nursery. (b) The corresponding scree plot showing the cumulative proportion of variance captured in the first 100 principal components with guides inserted where 25, 50, and 75 percent cumulative variance were captured (at components 3, 12, and 33, respectively).

Using this filtered marker set, a principal component analysis was applied to the marker data to assess any evident population structure. This was done by scaling the SNP matrix (X) with each SNP's mean to produce a centred marker matrix, Z. R's base `svd()` function was applied to Z to produce a matrix of left singular vectors and diagonal matrix of singular values (U and S, respectively) that were then used to derive the principal components of the original SNP covariance matrix (P = US). These components' scores along with the covariance matrix's eigenvalues were then used to assess population structure (Figure 3.1).

Lastly, this scaled SNP matrix, Z, was used to compute several genomic relationship matrices. Additive genomic relationship matrices were computed for the 456 parents (G_p) as well as on the 769 hybrids (G_h) using Van Raden's method ($G = \frac{ZZ^T}{2 \cdot \mathbf{p}^T(\mathbf{1}-\mathbf{p})}$)[41] via the R library, AGHmatrix[42]. The dominance relationship matrix on hybrids was also estimated using the method proposed by[43]. Unless explicitly stated otherwise, all marker-based analysis was carried out using the R statistical language[44].

Statistical Analysis

A two-stage modelling approach was conducted to first model within each trial to produce unbiased estimates for each hybrid (BLUEs) which in turn were used in multi-year and location prediction models.

Table 3.1: Selected spatial models across four field trials for percentage dry matter content (DM), total tuber number (TN), average tuber volume (TV), and total tuber yield (TY). Each entry is composed of the base model along with the variance structure of the residual.

	Est Field Trial		Heelsum Field Trial	
Trait	2019	2020	2019	2020
DM	(1) + (6)	(1) + (6)	(2) + (3)	(1) + (5)
TN	(1) + (6)	(1) + (6)	(1) + (6)	(1) + (6)
TV	(1) + (6)	(1) + (4)	(1) + (6)	(1) + (6)
TY	(1) + (6)	(1) + (6)	(1) + (6)	(1) + (6)

First-stage

For the first stage, a model selection procedure was conducted on each trial and trait to account for specific field trends and spatial dynamics. Eleven different spatial models were utilised across all 16 trait and trial combinations. From these eleven spatial models, only four were selected based upon the Akaike's Information Criterion (AIC); these four spatial models are described in detail here. The base model equation has the form:

$$y_{hb} = \mu + \alpha_h + \gamma_b + \varepsilon_{hb} \tag{3.1}$$

where the phenotype, y, is modelled on hybrid h (α) and block b (γ), both effects being treated as fixed. A residual, ε, was modelled as random but its variance structure varied per spatial model chosen. There was a modification of this base model structure where a random column effect was included:

$$y_{hbm} = \mu + \alpha_h + \gamma_b + c_m + \varepsilon_{hbm} \tag{3.2}$$

with an additional random effect ($c \sim \mathrm{N}(0, \sigma_c^2)$) on the plot in column m (c) in the trial field. There were four residual variance structures among the several spatial models chosen. The simplest was a residual variance with an independent and identically distributed (IID) Gaussian distribution such that:

$$\mathtt{var}(\varepsilon) = \sigma_e^2 \mathtt{I} \tag{3.3}$$

That is, the residual is modelled with a constant spatial variance (σ_e^2) which is invariant regardless of a plot's position in the field. The second variance structure utilizes a 1st order autoregressive structure along the column coordinates. This can be expressed as:

$$\mathtt{var}(\varepsilon) = \sigma_e^2 \mathtt{I}_r \otimes \Sigma_c(\rho_c) \tag{3.4}$$

where the residual variance is modelled as the Kronecker product of an identity matrix ($\mathtt{I}_r$: each dimension equal to the number of rows in trial field) and a 1$^{\text{st}}$ order autoregres-

sive matrix ($\mathtt{\Sigma}_c$: each dimension equal to number of columns in trial field) scaled by the spatial variance, σ_e^2. The autoregressive matrix, Σ_c, is parametrised by the autocorrelation parameter,ρ_c , so that the covariances decline exponentially (i.e., $\Sigma_{c_{ij}} = \rho_c^{|i-j|}$). The third residual structure builds on this with the addition of a spatially-independent error, often referred to as a *nugget effect*. This results in:

$$\mathtt{var}(\varepsilon) = \sigma_e^2 \mathtt{I}_r \otimes \Sigma_c(\rho_c) + \sigma_\eta^2 \tag{3.5}$$

where the spatially-independent variance, σ_η^2, is the error variance un-attributable to the spatial components, traditionally attributed to harvest & instrument related error[45]. The last spatial error builds on model (3.5) with the addition of an autoregressive effect along rows. This is then:

$$\mathtt{var}(\varepsilon) = \sigma_e^2 \Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c) + \sigma_\eta^2 \tag{3.6}$$

where Σ_r is a 1$^{\text{st}}$ order autoregressive matrix ($\mathtt{\Sigma}_r$: rank is same as number of rows in trial field) and defined by autocorrelation parameter, ρ_r. The combination of base model and residual variance structure is given in 3.1 for each trait and trial combination. BLUEs were then produced from each best supported spatial model and used for all genetic model training going forward.

Phenotypic variance & stability

To evaluate the genetic variance of the hybrids across trials, raw data from each trial was used to create the following multi-trial model:

$$y_{bfht} = \beta_f + \gamma_{fb} + a_{ht} + \varepsilon_{bfht} \tag{3.7}$$

where the hybrid response (y) is modelled on a fixed trial intercept (β) for trial, f, a fixed within-trial block effect (γ) for block b in trial f, a random hybrid by location interaction (a) for hybrid h in location t, and an IID Gaussian error term, ε ($\varepsilon \sim \mathrm{N}(0, \sigma_\varepsilon^2)$). The hybrid by location effect was modelled with an unstructured covariance structure so that there was a separate hybrid variance for each trial location along with their covariance. This is described by:

$$a \sim \mathrm{N}(\mathbf{0}, \Sigma_a) \qquad \Sigma_a = \begin{bmatrix} \sigma^2_{\mathrm{G(Est)}} & \sigma_{\mathrm{G(Est),G(Hee)}} \\ \sigma_{\mathrm{G(Hee),G(Est)}} & \sigma^2_{\mathrm{G(Hee)}} \end{bmatrix}$$

Using model (3.7), across-trial BLUPs were predicted for each hybrid. Several variance ratios were also produced from this model including broad-sense heritabilities ($\mathrm{H}^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_\varepsilon^2}$) per trial location and inter-genotype coefficients of variation ($\mathrm{CV_G} = \frac{\sigma_a}{\mu}$) using the variances embedded within Σ_a and means for each respective trial location. The intra-genotype coefficient of variation ($\mathrm{CV}_\varepsilon = \frac{\sigma_\varepsilon}{\mu}$) was also computed using the residual variance and global mean. These variance ratios are reported in Table 3.2 for each tuber phenotype.

All models used for first stage and phenotypic models were fit, examined, and tested using ASReml-R version 4 provided by VSN International[46].

Second-stage

Hybrid BLUEs from each trial were treated as the phenotypic response for second-stage models. BLUEs were carried along with their respective standard errors to properly weight each observation by its precision. This was done using a diagonal structure on the squared standard errors for each trial f to approximate the full residual covariance structure ($\mathrm{R} = \bigoplus_{f=1}^{t} \mathrm{R}_f$ with $\mathrm{R}_f = \mathrm{Diag}(\hat{\mathbf{s}}_f^2)$)[47,48]. These second-stage models take the form of a multi-environment model with two parametrisations for the hybrid genetic variance. Because we are interested in measuring the benefits of dominance genetic variance in genomic prediction, one model bases prediction solely on parental general combining ability effects (GCA model) and the second on both general and specific combining ability effects (GCA+SCA model). The GCA model is:

$$\hat{\mu}_{ijf} = \beta_f + g_i + g_j + \delta_{ij} + t_{ijf} + r_{ijf} \tag{3.8}$$

where $\hat{\mu}$, the hybrid BLUE, is modelled on a fixed trial intercept (β) for trial f, the GCA (g) of parents' i and j, a hybrid by trial interaction (t) with trial f and hybrid ij, a genetic residual term (δ) for hybrid ij, and a residual, r. All random effects have Gaussian distributions centred about 0 with the variances of GCA and hybrid-trial interactions being structured by the genomic relationship matrices, G_p, and G_h, respectively. GCA variance was modelled as a scaling of the additive parental genomic relationship matrix ($\sigma^2_{gca}\mathrm{G}_p$) while the hybrid by trial variance was represented by the Kronecker product of an identity matrix containing the GCA by trial interaction variance and the hybrid additive genomic relationship matrix ($\sigma^2_{gxt}\mathrm{I}_\mathrm{t} \otimes \mathrm{G}_h$). The variance of the genetic residual was modelled as IID so that the total hybrid variance according to the GCA model is $2\sigma^2_{gca}\mathrm{Z_p G_p Z_p^t} + \sigma^2_\delta \mathrm{I}$ where $\mathrm{Z_p}$ is the incidence matrix assigning parent lines to hybrid crosses. The residual variance is considered to be known (estimated in stage 1) so that the full residual distribution is $r_{ijf} \sim \mathrm{N}(0, \mathrm{R}_{ijf})$. The GCA+SCA model differs from the GCA model in the addition of the SCA in the hybrid making the full model:

$$\hat{\mu}_{ijf} = \beta_f + g_i + g_j + s_{ij} + \delta_{ij} + t_{ijf} + r_{ijf} \tag{3.9}$$

where s is the specific combining ability for parents' i and j. The variance of the SCA effect is defined by the dominance relationship matrix scaled by a scalar variance component ($\sigma^2_{sca}\mathrm{D}$) making the hybrid genetic variance in the GCA+SCA model equal to $2\sigma^2_{gca}\mathrm{Z_p G_p Z_p^t} + \sigma^2_{sca}\mathrm{D} + \sigma^2_\delta \mathrm{I}$. The modelling procedure of[32] was adapted by estimating the σ^2_{gca} and σ^2_{gxt} parameters in the GCA model followed by fixing these same components in the GCA+SCA models to these estimates. These model constraints allowed for partitioning of the GCA and SCA variance. Having said this, the genetic residual term in addition to containing higher order interactions, is not orthogonal with the SCA term. This is partially reflected in deviations in total genetic variance between GCA and GCA+SCA models (Table 3.3).

Table 3.2: Variance ratios from multi-trial model for percentage dry matter content (DM), total tuber number (TN), average tuber volume (TV), and total tuber yield (TY). The ratios include broad-sense heritabilities observed in Est and Heelsum ($H_{Est}{}^2$ & $H_{Hee}{}^2$), the inter-genotype coefficients of variation from Est and Heelsum ($CV_{G:Est}$ & $CV_{G:Hee}$) & the intra-genotype coefficients of variation CV_{ε}.

	DM	TN	TV	TY
H^2_{Est}	0.68	0.49	0.73	0.47
H^2_{Hee}	0.55	0.73	0.82	0.81
$CV_{G(Est)}$	11.06	23.57	24.82	28.33
$CV_{G(Hee)}$	9.19	27.77	27.81	35.91
CV_{ε}	7.91	19.74	14.00	22.11

Model Testing

To test the advantage of including dominance effects in GP for hybrid potato, we examine the performance of the GCA (3.8) and GCA+SCA (3.9) models. The authors define model performance by the prediction accuracy and error on a given test set. Prediction accuracy was assessed using Pearson's correlation coefficient between true and predicted hybrid performance ($\rho_{y,\hat{y}} = \frac{\text{cov}(y,\hat{y})}{\text{sd}(y)\cdot\text{sd}(\hat{y})}$) and prediction error by the root mean square error scaled by the trait mean for convenience (Scaled RMSE $= 100 \cdot \frac{\sqrt{\text{E}((y-\hat{y})^2)}}{\mu}$). The testing and training strategy was conducted by defining two test sets with different levels of parental information shared between training and test sets[49]. This procedure begins by randomly assigning all hybrid parents into an evaluated and unevaluated set. Consequently, these parental groups randomly assign hybrids to three sets, that is, hybrids with zero evaluated parents (0EP), hybrids with one evaluated parent (1EP), and hybrids with two evaluated parents (the training set). This allows for model testing which mirrors breeding scenarios where a hybrid is produced from a test cross of elite and novel parents (1EP) and crossing of two novel parents (0EP). Because one of the aims of this study was evaluating prediction of genotype by trial effects, we define true hybrid performance as (1) the hybrid average performance across trials (AVG) and as (2) performance within individual trials (E19, E20, H19, H20). This testing scheme was repeated 100 times for each tuber phenotype and genetic model to assess model accuracy (Figure 3.2) and error (Figure 3.3) under multiple scenarios. Because absorption of coefficient matrices failed using ASReml-R, model fitting and cross-validation for second stage models (3.8 and 3.9) were instead performed using the sommer R package (https://github.com/covaruber/sommergithub.com/covaruber/sommer), a multivariate mixed model solver, using its average information method[50].

3.3 Results

Marker Analysis

The filtered SNP dataset was used for population analysis and the construction of both additive genomic relationship matrices (G_p and G_h) and the dominance relationship matrix for hybrids (D). The principal component analysis revealed little evident population structure aside from a few minor family clusters (Figure 3.1 (a)). Less than 20% of the original marker covariance was captured on the first two principal axes with 75% of its variance distributed across the first 33 axes (Figure 3.1 (b)). Along with this, there was no clear relationship between the older and newer nurseries and the first or second principal loadings suggesting little differentiation between these different sub-populations over time. This analysis was complemented with the Bayesian population software, STRUCTURE, to determine the number of ancestral founders that gave rise to this population[51]. This analysis found no evidence of recent admixture events with one ancestral population being the most probable origin for this panel of inbred lines. Both findings suggest little population structure in this panel needing to be taken into account whilst modelling.

Trial & Phenotype Analysis

Among all 16 trial and trait combinations, the majority were best fit with both an autoregressive trend and nugget effect (Table 3.1). However, a variant of a traditional RCBD design was selected for dry matter content in Heelsum 2019. The severity of the spatial heterogeneity varied per trait in each trial, but spatial trends in the Est trial field in 2019 were large for all phenotypes. The spatial model which assumed an IID residual (i.e. dry matter content in Heelsum 2019) had empirical semivariograms which were relatively stable with exception of the last 60 observations consistent with an edge effect near the end of the trial field.

Using the variance components from model 3.7, multiple variance ratios were produced for all tuber phenotypes (Table 3.2). Broad-sense heritabilities ranged between 0.47-0.73 in Est and 0.55-0.82 in Heelsum with average tuber volume being the most heritable trait over both locations. Heritabilities between Est and Heelsum were most divergent for total yield (0.47 and 0.81, respectively) and tuber number (0.49 and 0.73, respectively). This coincides with the genetic coefficients of variation which exhibited greater hybrid variation in Heelsum (between 27% and 35%) than the Est trial field (between 23% and 28%) for all traits but dry matter content. Dry matter content showed a larger genetic variance in Est over Heelsum (H^2 of 0.68 versus 0.55), in contrast to what was seen in the other phenotypes. This could be due to a positive bias in tuber dry matter content among smaller tubers which were more prevalent under the clay conditions of the Est field trial. The error coefficient of variation was similar among all yield phenotypes (between 14% in average tuber volume and 22% in total tuber yield) with dry matter content having the smallest percentage of error (7.9%). The small coefficients of variation seen in dry matter content (relative to other traits) are somewhat artefactual due to its scale. The maximum variance that can be observed in dry matter content is much smaller due to the very tight bounds around biologically-relevant dry matter values (See (Bhatia Davis,

2000)[52]).

Table 3.3: Variance components for the GCA and GCA+SCA genetic models for percentage dry matter content (DM), total tuber number (TN), average tuber volume (TV), and total tuber yield (TY). Included are the variances attributed to general combining ability, hybrid by trial interaction, specific combining ability, genetic residual, and the proportions of genetic variance among each genetic component.

Trait	Variance Components:	σ^2_{gca}	σ^2_{gxt}	σ^2_{sca}	σ^2_{δ}	$\frac{2\sigma^2_{gca}}{\sigma^2_G}$	$\frac{\sigma^2_{sca}}{\sigma^2_G}$	$\frac{\sigma^2_{\delta}}{\sigma^2_G}$
DM	GCA	1.1	0.5		0.4	0.84		0.16
	GCA+SCA			0.3	0.3	0.80	0.09	0.11
TN	GCA	368.9	287.7		341.0	0.68		0.32
	GCA+SCA			195.2	232.2	0.63	0.17	0.20
TV	GCA	11.8	2.7		7.2	0.76		0.24
	GCA+SCA			5.5	4.4	0.71	0.16	0.13
TY	GCA	7.4	7.8		5.3	0.74		0.26
	GCA+SCA			3.0	3.6	0.69	0.14	0.17

genetic-modelling-genomic-prediction

Genetic Modelling & Genomic Prediction

Both the GCA (3.8) and GCA + SCA (3.9) sets of models converged for all traits studied here. Credible variance components were also extracted for all eight models (Table 3.3). GCA variance was the largest variance component for all traits except total tuber yield where GCA variance was nearly equal with hybrid by trial interaction component (TY: $\sigma^2_{gca} = 7.4$, $\sigma^2_{gxt} = 7.8$). With respect to total genetic variance, the additive variance was the largest genetic effect for all phenotypes (between 63% and 84%). Proportion of SCA genetic variance in the full GCA+SCA models was noticeably smaller in dry matter content (9%) in contrast to the other yield components (14-17%). It's worth noting that the genetic residual term in both the GCA and GCA+SCA models contained a considerable portion of genetic variance. This was most apparent in both GCA and GCA+SCA models for tuber number (32% and 20%, respectively) though present in total tuber yield (26% and 17%, respectively) and average tuber volume (24% and 13%, respectively) as well.

Turning to model performance, there were no measurable difference between GCA and GCA+SCA models in predicting average hybrid performance (Table 3.4) or among any of the trial-specific prediction targets (Figure 3.2). As expected, prediction accuracies were globally deflated in the 0EP set ($\rho_{y,\hat{y}}$ between 0.36 and 0.61) in contrast to the 1EP set ($\rho_{y,\hat{y}}$ between 0.50 and 0.68) (Table 3.4). These coincided with larger root mean square errors in the 0EP set (scaled RMSE = $8\% - 26.4\%$) over the 1EP set (scaled RMSE = $7.4\% - 24.5\%$). Looking among traits, prediction accuracy was highest in average tuber

Table 3.4: Median prediction accuracies and median scaled root mean square error for each trait, genetic model, and test set. This was conducted on dry matter content (DM), total tuber number (TN), mean tuber volume (TV), and total tuber yield (TY). Predictions were tested against average hybrid performance (AVG) produced from model 3.7.

	0 Evaluated Parents Set		1 Evaluated Parents Set	
Phenotype	GCA	GCA+SCA	GCA	GCA+SCA
DM	0.49 (8.0)	0.49 (8.0)	0.58 (7.4)	0.58 (7.4)
TN	0.36 (20.3)	0.36 (20.2)	0.50 (19.2)	0.51 (19.2)
TV	0.61 (20.7)	0.61 (20.7)	0.68 (19.0)	0.68 (18.9)
TY	0.46 (26.4)	0.46 (26.4)	0.58 (24.5)	0.58 (24.5)

volume (0EP: 0.61 and 1EP: 0.68) followed by dry matter content (0EP: 0.49 and 1EP: 0.58) and total tuber yield (0EP: 0.46 and 1EP: 0.58) with considerably lower prediction error in the models for dry matter (7.4% − 8.0%) than those in average tuber volume (18.9% − 20.7%) and total tuber yield (24.5% − 26.4%). Tuber number had the lowest accuracy among traits ($\rho_{y,\hat{y}}$ between 0.36 and 0.51) with intermediate prediction errors (18.9%−20.7%). When examining performance among the trial-specific targets, prediction accuracies do vary per trait and trial-target. In general, prediction accuracies tended to be lower among the 2019 trial sets (i.e. E19 and H19) for dry matter content and average tuber volume while accuracy were lower in the 2020 trial sets (i.e. E20 and H20) for tuber number and total tuber yield. These differences in performance between trial targets, however, become much more apparent while examining the prediction error (Figure 3.3). The root mean square error was largest (and most variable among prediction targets) in total yield, and to a lesser extent, tuber number, with errors being highest for the Heelsum trial sets (i.e. H19 and H20) in both phenotypes. Average tuber volume also exhibited a higher RMSE for predictions in the Heelsum 2020 target while the RMSE's in dry matter content were much smaller and relatively uniform across trait targets.

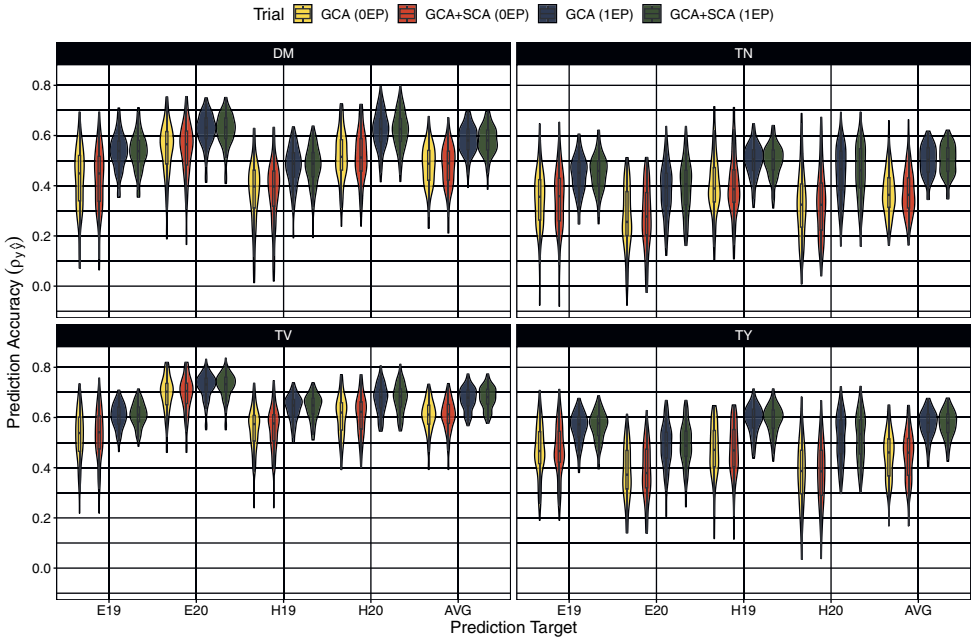


Figure 3.2: Predictive accuracy (correlation coefficient of true hybrid performance and predicted cross value) in the zero evaluated parent set (0EP) and the one evaluated parent set (1EP) for GCA model and GCA+SCA models for dry matter content (DM), total tuber number (TN), mean tuber volume (TV), and total tuber yield (TY). Each trait, model, & testing scenario were tested 100 times. Predictions were either compared to BLUEs generated from the first stage of modelling for Est 2019 and 2020 (E19 and E20, respectively) and Heelsum 2019 and 2022 (H19 and H20, respectively) or against across-trial BLUPs (AVG) produced from model (3.7).

3.4 Discussion

Hybrid breeding has been revolutionized by GS and GP making adoption of these techniques a key step in accelerating quantitative improvement. Of particular importance when implementing these technologies is determining the contribution and nature of the genetic effects in the relevant target and then assessing how molecular marker and pedigree information can best provide predictive power for it. This in turn provides a roadmap for breeders to place these predictive methods where they can be most efficiently applied. Additionally, understanding the presence of genotype by environment interaction has also received increasing attention as selection continues to shift away from general performance to adaptability to specific envirotypes. Our findings represent a first inspection of GP for multiple tuber traits in hybrid potato using conventional genetic models.

Feasibility of genomic prediction in hybrid potato

GP was possible in both testing scenarios (0EP and 1EP) in both genetic models (GCA and GCA+SCA) for all tuber traits, though not equally among them. Average tuber volume had the highest prediction accuracy in both test scenarios (0.61 & 0.68, respectively) followed by dry matter content and total tuber yield. Looking among other available studies, dry matter content (and starch) has consistently high prediction accuracies (0.73-0.84) in tetraploid populations[27,30,34]. Total yield (and components of yield) were more variable (0.31 - 0.66) reflecting different modelling strategies and training set properties[30,33]. Among our traits studied, total tuber number exhibited the poorest prediction accuracy especially in the 0EP testing scenario (Figure 3.4).[33] similarly found relatively poor model performance for tuber number with their additive model showing an average accuracy of 0.35. In contrast to the tetraploid studies, our prediction accuracies were notably lower. This can likely be attributed to differences in testing strategies and a smaller training to test set ratio (expectation of 1 and 0.5 for the 0EP and 1EP sets, respectively) resulting in a more severe testing in contrast to the K-fold cross-validation strategy invoked elsewhere.

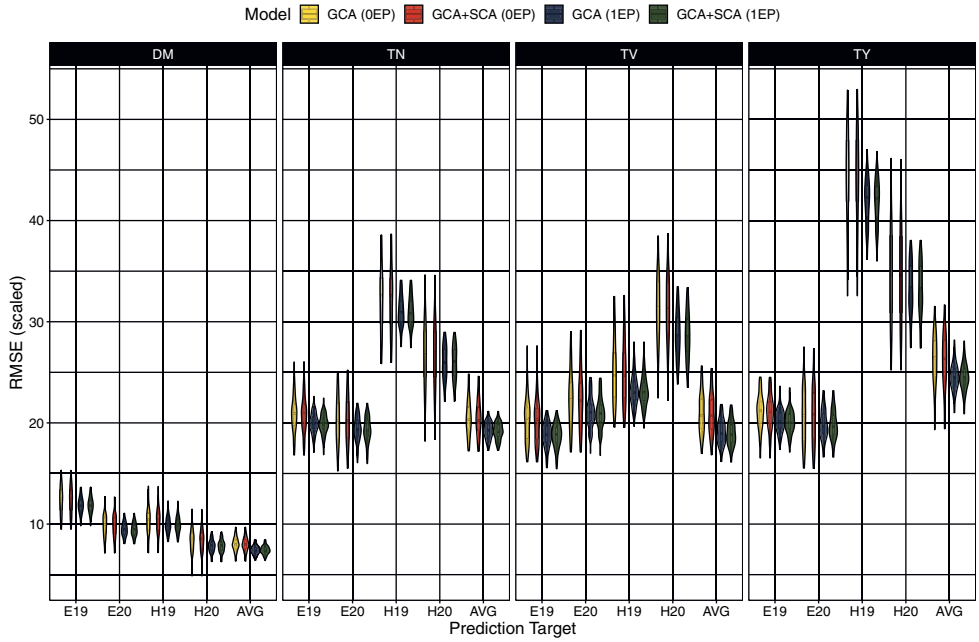


Figure 3.3: Scaled root mean square error ($100 \cdot \sqrt{\mathsf{E}(\hat{y} - y)^2}/\mu$) in prediction in the zero common parent (0EP) and one common parent (1EP) sets for GCA and GCA+SCA models for dry matter content (DM), total tuber number (TN), mean tuber volume (TV), and total tuber yield (TY). Each trait, model, & testing scenario were tested 100 times. Predictions were either compared to BLUEs generated from the first stage of modelling for Est 2019 and 2020 (E19 and E20, respectively) and Heelsum 2019 and 2022 (H19 and H20, respectively) or against across-trial BLUPs (AVG) produced from model 3.7.

Contribution of non-additive effects in prediction

We found no practical benefit in the incorporation of dominance effects (via SCA) for GP model performance in any tuber phenotype. The lack of added value with SCA also appears inconsistent given the magnitude of SCA variances measured here (Table 3.3) and in our preceding phenotypic study[36]. As highlighted there, several lines of reasoning can be pursued while considering non-additive genetic effects. Reflecting on the sparse crossing design used to generate these hybrids, its not unexpected for shrinkage of the SCA effects even if the model is technically identifiable. If this is the case, then larger training sets would be required for exploitation of SCA among hybrids. Alternatively, reflecting on the size of the genetic residual (between 20 and 37% of σ^2_G), kernel-based approaches could be more informative in predicting hybrid performance[53,54]. Rather than decomposing the total genetic value into separate additive and non-additive components, prediction is instead based upon the total observed genetic value[55]. The application of such semi-parametric approaches could be an interesting avenue for future prediction modelling in DHP.

Hybrid prediction and genotype by environment interaction

While present, genotype by environment interactions did not appear to affect model performance for dry matter content and average tuber volume. This coincides with relatively consistent heritabilities between trial locations and the small σ^2_{gxt} variance components for both traits. Several studies in diploid[38] and tetraploid[30,32,34,56] potato also corroborate the high stability of dry matter content. Examining total yield and tuber number, prediction accuracies appear fairly stable across target environments, but prediction errors were highly dependent on target environment (Figure 3.3). This was especially apparent between Heelsum and Est 2019 trials where the difference in median scaled RMSE was over 25% and 12%, respectively. Considering the significant differences in heritabilities between trial locations (Table 3.2), this could negatively impact prediction accuracy if the genotype by environment structure does not reflect this. Similar to what has been reported in tetraploid[30,34,56], total yield is highly influenced by genotype by environment interaction and difficult to structure. For future work in genotype by environment interactions, designing trials around known abiotic stresses in potato would allow for more meaningful decomposition of these effects. Doing so in climates which target relevant condition for the world's potato growers is also important in a GP context for prediction of hybrid adaptability.

Genomic prediction for breeders

These results have some implications for potato breeders. First, this study gives ample evidence that stable additive genetic effects are not only able to be estimated for potato inbred lines, but can also be leveraged for the prediction of a parent's breeding value in a dedicated cross. In particular, these results show the potential of genomic prediction in both a test cross and novel crossing set context through the use of the 1EP and 0EP test sets, respectively. Based upon these results, confident predictions can be offered in both contexts for all phenotypes studied here with the exception of total tuber number which

exhibited depressed prediction accuracies making them too unreliable in a novel crossing set for breeders. The GCA variance being the largest genetic effect among all traits is also worth reiterating as it directs breeders towards prioritizing additive variance among multiple tuber traits. This directly touches upon a breeding program's structure and how it exploits additive variation during population improvement among highly heritable traits.

3.5 Conclusions

In this study, we have laid out the groundwork for genomic prediction in hybrid potato with simple genetic models and no training set optimization. We demonstrate the GEBVs estimated from an additive model (our GCA model) suffices in predicting hybrid potato performance in dry matter content, average tuber volume, and total yield. The effect of genotype by environment interaction is mostly relevant for yield components like tuber number and total yield while dry matter content and average tuber volume were remarkably stable among target environments for genomic prediction. Principally, these results show for the first time that genomic prediction can be leveraged for inbred and hybrid selection in diploid hybrid potato. The adoption of predictive breeding in diploid potato is invaluable for the continued genetic progress of this new hybrid crop.

Funding

The research produced for this publication was made possible through funding provided by Topsector Tuinbouw & Uitgangsmaterialen in the 'Re-booting potato: enhancing the breeding of hybrid diploid potato' project (project number: TU-18155).

Bibliography

1. Adams J, De Vries M, and Van Eeuwijk F. Efficient Genomic Prediction of Yield and Dry Matter in Hybrid Potato. Plants. 2023 Jul 11; 12:2617. DOI: 10.3390/plants12142617. Available from: https://www.mdpi.com/2223-7747/12/14/2617
2. FAO.STAT. Food and Agriculture Organization of the United Natons. FAOSTAT database, 2023. Available from: http://www.fao.org/faostat [Accessed on: 2023 Jun 5]
3. Haverkort AJ. Ecology of Potato Cropping Systems in Relation to Latitude and Altitude. Agricultural Systems. 1990 Jan; 32:251–72. DOI: 10.1016/0308-521X(90)90004-A. Available from: https://www.sciencedirect.com/science/article/pii/0308521X9090004A [Accessed on: 2023 May 6]
4. Douches DS, Maas D, Science KJC, and 1996 U. Assessment of Potato Breeding Progress in the USA over the Last Century. Crop Science. 1996; 36:1544–52. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/36/6/CS0360061544 [Accessed on: 2019 Apr 29]
5. Lian Q, Tang D, Bai Z, Qi J, Lu F, Huang S, and Zhang C. Acquisition of Deleterious Mutations during Potato Polyploidization. Journal of Integrative Plant Biology. 2019 Jan 1; 61:7–11. DOI: 10.1111/jipb.12748
6. Gebhardt C. Bridging the Gap between Genome Analysis and Precision Breeding in Potato. Trends in Genetics. 2013 Apr; 29:248–56. DOI: 10.1016/j.tig.2012.11.006. Available from: https://www.sciencedirect.com/science/article/pii/S0168952512001904 [Accessed on: 2023 Feb 1]
7. Gopal J. Challenges and Way-forward in Selection of Superior Parents, Crosses and Clones in Potato Breeding. Potato Research. 2015 Jun; 58:165–88. DOI: 10.1007/s11540-015-9292-6. Available from: https://link-springer-com.ezproxy.library.wur.nl/article/10.1007/s11540-015-9292-6 [Accessed on: 2023 Jan 20]
8. Bradshaw JE. A Brief History of the Impact of Potato Genetics on the Breeding of Tetraploid Potato Cultivars for Tuber Propagation. Potato Research. 2022 Sep; 65:461–501. DOI: 10.1007/s11540-021-09517-w. Available from: https://link.springer.com/article/10.1007/s11540-021-09517-w [Accessed on: 2023 May 4]
9. Lindhout P, Meijer D, Schotte T, Hutten RC, Visser RG, and Eck HJ van. Towards F1 hybrid seed potato breeding. Potato Research 2011; 54:301–12
10. Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS, Bamberg JB, Thompson AL, Biz-

imungu B, Holm DG, Brown CR, Haynes KG, Sathuvalli VR, Veilleux RE, Creighton Miller J, Bradeen JM, and Jiang J. Reinventing Potato as a Diploid Inbred Line-Based Crop. Crop Science. 2016 Jul 1; 56:1412–22. DOI: 10.2135/cropsci2015.12.0740

11. Van Dijk L, de Vries M, Lommen W, and Struik P. Transplanting Hybrid Potato Seedlings at Increased Densities Enhances Tuber Yield and Shifts Tuber-Size Distributions. Potato Research. 2021. DOI: 10.1007/s11540-021-09522-z
12. Technow F, Podlich D, and Cooper M. Back to the Future: Implications of Genetic Complexity for the Structure of Hybrid Breeding Programs. G3 Genes\textbarGenomes\textbarGenetics. 2021 Jul; 11:jkab153. DOI: 10.1093/g3journal/jkab153. Available from: https://doi.org/10.1093/g3journal/jkab153 [Accessed on: 2023 May 4]
13. Technow F, Riedelsheimer C, Schrag TA, and Melchinger AE. Genomic Prediction of Hybrid Performance in Maize with Models Incorporating Dominance and Population Specific Marker Effects. Theoretical and Applied Genetics. 2012 Oct; 125:1181–94. DOI: 10.1007/s00122-012-1905-8. Available from: https://doi.org/10.1007/s00122-012-1905-8 [Accessed on: 2023 May 6]
14. Pallais N. True Potato Seed: Changing Potato Propagation from Vegetative to Sexual. HortScience. 1991 Mar; 26:239–41
15. Eggers EJ, van der Burgt A, van Heusden SAW, de Vries ME, Visser RGF, Bachem CWB, and Lindhout P. Neofunctionalisation of the Sli Gene Leads to Self-Compatibility and Facilitates Precision Breeding in Potato. Nature Communications. 2021 Jul; 12:4141. DOI: 10.1038/s41467-021-24267-6. Available from: https://www.nature.com/articles/s41467-021-24267-6 [Accessed on: 2023 Jan 20]
16. Ma L, Zhang C, Zhang B, Tang F, Li F, Liao Q, Tang D, Peng Z, Jia Y, Gao M, Guo H, Zhang J, Luo X, Yang H, Gao D, Lucas WJ, Li C, Huang S, and Shang Y. A nonS-locus F-box Gene Breaks Self-Incompatibility in Diploid Potatoes. Nature Communications. 2021 Jul; 12:4142. DOI: 10.1038/s41467-021-24266-7. Available from: https://www.nature.com/articles/s41467-021-24266-7 [Accessed on: 2023 Feb 1]
17. Lindhout P, de Vries M, ter Maat M, Ying S, Viquez-Zamora Marcela, and van Deusden S. Hybrid Potato Breeding for Improved Varieties. *Achieving Sustainable Cultivation of Potatoes Volume 1*. Ed. by Wang-Pruski Gefu. Burleigh Dodds Science Publishing, 2018 Apr 9:119–42. DOI: 10.1201/9781351114455-10/HYBRID-POTATO-BREEDING-IMPROVED-VARIETIES-GEFU-WANG-PRUSKI
18. Bethke PC, Halterman DA, Francis DM, Jiang J, Douches DS, Charkowski AO, and Parsons J. Diploid Potatoes as a Catalyst for Change in the Potato Industry. American Journal of Potato Research. 2022 Dec; 99:337–57. DOI: 10.1007/s12230-022-09888-x. Available from: https://doi.org/10.1007/s12230-022-09888-x [Accessed on: 2023 Feb 1]
19. Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, and Huang S. Genome Design of Hybrid Potato. Cell. 2021 Jul 22; 184:3873–3883.e12. DOI: 10.1016/J.CELL.2021.06.006

20. Labroo MR, Studer AJ, and Rutkoski JE. Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. Frontiers in Genetics. 2021 Feb 24; 12:234. DOI: 10.3389/FGENE.2021.643761
21. Zhao Y, Mette MF, and Reif JC. Genomic Selection in Hybrid Breeding. Plant Breeding. 2015; 134:1–10. DOI: 10.1111/pbr.12231. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/pbr.12231 [Accessed on: 2023 May 1]
22. Meuwissen THE, Hayes BJ, and Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics. 2001 Apr 1; 157:1819–29. DOI: 10.1093/genetics/157.4.1819. Available from: https://academic.oup.com/genetics/article/157/4/1819/6048353 [Accessed on: 2022 Nov 13]
23. Bernardo R. Bandwagons I, Too, Have Known. Theoretical and Applied Genetics. 2016 Dec; 129:2323–32. DOI: 10.1007/s00122-016-2772-5. PMID: 27681088
24. Rembe M, Zhao Y, Jiang Y, and Reif JC. Reciprocal Recurrent Genomic Selection: An Attractive Tool to Leverage Hybrid Wheat Breeding. Theoretical and Applied Genetics. 2019 Mar; 132:687–98. DOI: 10.1007/s00122-018-3244-x. Available from: https://link.springer.com/article/10.1007/s00122-018-3244-x [Accessed on: 2023 May 6]
25. Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Mackay IJ, and Hickey JM. A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. Crop Science. 2017; 57:2372–86. DOI: 10.2135/cropsci2016.09.0742. Available from: https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci2016.09.0742 [Accessed on: 2023 May 4]
26. Heffner EL, Lorenz AJ, Jannink JL, and Sorrells ME. Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Science. 2010 Sep; 50:1681–90. DOI: 10.2135/cropsci2009.11.0662. Available from: https://onlinelibrary.wiley.com/doi/10.2135/cropsci2009.11.0662 [Accessed on: 2023 May 6]
27. Sverrisdóttir E, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, Janss L, Bryan G, and Nielsen KL. The Value of Expanding the Training Population to Improve Genomic Selection Models in Tetraploid Potato. Frontiers in Plant Science. 2018; 9. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2018.01118 [Accessed on: 2022 Dec 9]
28. Pandey J, Scheuring DC, Koym JW, Endelman JB, and Vales MI. Genomic Selection and Genome-Wide Association Studies in Tetraploid Chipping Potatoes. The Plant Genome. 2022 May; n/a:e20297. DOI: 10.1002/tpg2.20297. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/tpg2.20297 [Accessed on: 2023 Feb 1]
29. Enciso-Rodriguez F, Douches D, Lopez-Cruz M, Coombs J, and de los Campos G. Genomic Selection for Late Blight and Common Scab Resistance in Tetraploid Potato (Solanum Tuberosum). G3 Genes\textbarGenomes\textbarGenetics. 2018 Jul; 8:2471–81. DOI: 10.1534/g3.118.200273. Available from: https://doi.org/10.1534/g3.118.200273 [Accessed on: 2023 Apr 25]
30. Ortiz R, Crossa J, Reslow F, Perez-Rodriguez P, and Cuevas J. Genome-Based Genotype × Environment Prediction Enhances Potato (Solanum Tuberosum L.)

Improvement Using Pseudo-Diploid and Polysomic Tetraploid Modeling. Frontiers in Plant Science. 2022; 13. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2022.785196 [Accessed on: 2023 Apr 25]

31. Byrne S, Meade F, Mesiti F, Griffin D, Kennedy C, and Milbourne D. Genome-Wide Association and Genomic Prediction for Fry Color in Potato. Agronomy. 2020 Jan; 10:90. DOI: 10.3390/agronomy10010090. Available from: https://www.mdpi.com/2073-4395/10/1/90 [Accessed on: 2023 Feb 1]
32. Endelman JB, Carley CA, Bethke PC, Coombs JJ, Clough ME, da Silva WL, De Jong WS, Douches DS, Frederick CM, Haynes KG, Holm DG, Miller JC, Muñoz PR, Navarro FM, Novy RG, Palta JP, Porter GA, Rak KT, Sathuvalli VR, Thompson AL, and Yencho GC. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics. 2018 May 1; 209:77–87. DOI: 10.1534/genetics.118.300685. PMID: 29514860. Available from: https://pubmed.ncbi.nlm.nih.gov/29514860/ [Accessed on: 2021 Jan 29]
33. Wilson S, Malosetti M, Maliepaard C, Mulder HA, Visser RGF, and van Eeuwijk F. Training Set Construction for Genomic Prediction in Auto-Tetraploids: An Example in Potato. Frontiers in Plant Science. 2021; 12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.771075 [Accessed on: 2022 Nov 15]
34. Cuevas J, Reslow F, Crossa J, and Ortiz R. Modeling Genotype × Environment Interaction for Single- and Multi-Trait Genomic Prediction in Potato (*Solanum* Tuberosum L.) preprint. Plant Biology, 2022 Aug. DOI: 10.1101/2022.08.09.503418. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.08.09.503418 [Accessed on: 2022 Nov 12]
35. Slater AT, Cogan NO, Forster JW, Hayes BJ, and Daetwyler HD. Improving Genetic Gain with Genomic Selection in Autotetraploid Potato. The Plant Genome. 2016; 9:plantgenome2016.02.0021. DOI: 10.3835/plantgenome2016.02.0021. Available from: https://onlinelibrary.wiley.com/doi/abs/10.3835/plantgenome2016.02.0021 [Accessed on: 2023 Jan 20]
36. Adams JR, de Vries ME, Zheng C, and van Eeuwijk FA. Little Heterosis Found in Diploid Hybrid Potato: The Genetic Underpinnings of a New Hybrid Crop. G3: Genes, Genomes, Genetics. 2022 Jun; 12. DOI: 10.1093/g3journal/jkac076. PMID: 35460241
37. Wang X, Zhang Z, Xu Y, Li P, Zhang X, and Xu C. Using Genomic Data to Improve the Estimation of General Combining Ability Based on Sparse Partial Diallel Cross Designs in Maize. Crop Journal. 2020 Oct 1; 8:819–29. DOI: 10.1016/j.cj.2020.04.012
38. Stockem J, de Vries M, van Nieuwenhuizen E, Lindhout P, and Struik PC. Contribution and Stability of Yield Components of Diploid Hybrid Potato. Potato Research. 2020 Sep 1; 63:345–66. DOI: 10.1007/s11540-019-09444-x. Available from: https://doi.org/10.1007/s11540-019-09444-x [Accessed on: 2021 Feb 2]
39. LGC. SeqSNP Targeted GBS as Alternative for Array Genotyping in Routine Breeding Programs [White Paper]. Biosearch Technologies, 2019. Available from: https:

//biosearch-cdn.azureedge.net/assetsv6/seqsnp-tgbs-alternative-genotyping-routine-breeding-programs.pdf

40. Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, and Cierco-Ayrolles C. Novel Measures of Linkage Disequilibrium That Correct the Bias Due to Population Structure and Relatedness. Heredity. 2012 Mar; 108:285–91. DOI: 10.1038/hdy.2011.73. PMID: 21878986
41. VanRaden PM. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science. 2008 Nov; 91:4414–23. DOI: 10.3168/jds.2007-0980. PMID: 18946147
42. Amadeu RR, Cellon C, Olmstead JW, Garcia AAF, Resende MFR, and Muñoz PR. AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example. The Plant Genome. 2016 Nov; 9. DOI: 10.3835/plantgenome2016.01.0009. Available from: https://onlinelibrary.wiley.com/doi/10.3835/plantgenome2016.01.0009 [Accessed on: 2022 Nov 8]
43. Su G, Christensen OF, Ostersen T, Henryon M, and Lund MS. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. PLOS ONE. 2012 Sep; 7:e45293. DOI: 10.1371/journal.pone.0045293. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045293 [Accessed on: 2023 Mar 2]
44. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2022. Available from: https://www.R-project.org/
45. Müller BU, Kleinknecht K, Möhring J, and Piepho HP. Comparison of Spatial Models for Sugar Beet and Barley Trials. Crop Science. 2010 May; 50:794–802. DOI: 10.2135/cropsci2009.03.0153
46. Butler DG, Cullis BR, Gilmour AR, Gogel BJ, and Thompson R. ASReml-R Reference Manual Version 4. Hemel Hempstead, HP1 1ES: VSN International Ltd, 2017. Available from: http://www.homepages.ed.ac.uk/iwhite/asreml/uop. [Accessed on: 2021 Jan 19]
47. Frensham A, Cullis B, and Verbyla A. Genotype by Environment Variance Heterogeneity in a Two-Stage Analysis. Biometrics. 1997 Dec; 53:1373. DOI: 10.2307/2533504
48. Möhring J and Piepho HP. Comparison of Weighting Methods in Two-Stage Analysis of Plant Breeding Trials. Crop Science - CROP SCI. 2009 Nov; 49. DOI: 10.2135/cropsci2009.02.0083
49. Schrag TA, Möhring J, Maurer HP, Dhillon BS, Melchinger AE, Piepho HP, Sørensen AP, and Frisch M. Molecular Marker-Based Prediction of Hybrid Performance in Maize Using Unbalanced Data from Multiple Experiments with Factorial Crosses. Theoretical and Applied Genetics. 2009 Feb; 118:741–51. DOI: 10.1007/s00122-008-0934-9. PMID: 19048224
50. Covarrubias-Pazaran G. Genome-Assisted Prediction of Quantitative Traits Using the R Package Sommer. PLOS ONE. 2016 Jun; 11:e0156744. DOI: 10.1371/journal.pone.0156744. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0156744 [Accessed on: 2023 Jan 25]

51. Falush D, Stephens M, and Pritchard JK. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics. 2003 Aug; 164:1567–87. DOI: 10.1093/genetics/164.4.1567. Available from: https://academic.oup.com/genetics/article/164/4/1567/6050225 [Accessed on: 2022 Nov 12]
52. Bhatia R and Davis C. A Better Bound on the Variance. The American Mathematical Monthly. 2000 Apr; 107:353–7. DOI: 10.1080/00029890.2000.12005203. Available from: https://doi.org/10.1080/00029890.2000.12005203 [Accessed on: 2023 May 9]
53. Gianola D and van Kaam JBCHM. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics. 2008 Apr; 178:2289–303. DOI: 10.1534/genetics.107.084285. Available from: https://doi.org/10.1534/genetics.107.084285 [Accessed on: 2022 Dec 9]
54. Crossa J, Martini JW, Gianola D, Pérez-Rodríguez P, Jarquin D, Juliana P, Montesinos-López O, and Cuevas J. Deep Kernel and Deep Learning for Genome-Based Prediction of Single Traits in Multienvironment Breeding Trials. Frontiers in Genetics. 2019 Dec; 10:1168. DOI: 10.3389/fgene.2019.01168. PMID: 31921277. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913188/ [Accessed on: 2023 Apr 25]
55. Bernardo R. Reinventing Quantitative Genetics for Plant Breeding: Something Old, Something New, Something Borrowed, Something BLUE. Heredity. 2020 Apr 15 :1–11. DOI: 10.1038/s41437-020-0312-1. Available from: https://doi.org/10.1038/s41437-020-0312-1 [Accessed on: 2020 Jul 9]
56. Wilson SE. Statistical Considerations for Applying Genomic Prediction to Potato. PhD thesis. Wageningen University, 2023 Feb. DOI: 10.18174/583098. Available from: https://research.wur.nl/en/publications/300f1ea4-aa7b-462d-94be-8e05b582ce3e [Accessed on: 2023 Mar 27]

4

Digging up QTL: Evaluating the efficiency of marker-assisted selection in hybrid potato

Abstract

Hybrid potato is a novel crop strategy which develops cultivars with improved F1 hybrids created from inbred diploid potato lines. As with other crops, the hybrid breeding system allows for greater flexibility and potentially greater speed in contrast to traditional tetraploid potato breeding. An important aspect of how a hybrid breeding program should be structured is resource and technology allocation for optimum selection efficiency. The authors of this paper examined this question with respect to marker and selection technologies examining the relative efficiency of marker-assisted selection (`MA`) relative to a traditional genome-wide selection (`GW`) for F4 candidate selection. For the MA scenario, we were able to identify 35 QTL for all tuber phenotypes that accounted for a large portion of the genetic variation in multiple yield components. The `GW` strategy had the highest prediction accuracy with relative selection efficiency's consistent with this for all traits. However, when examining the relative efficiency with respect to cost of genotyping strategy, the `MA` scenario was more efficient for all traits studied here. Our findings suggest that marker-assisted selection could be an attractive method for quantitative trait selection in commercial hybrid potato programs.

4.1 Introduction

Potato (*Solanum tuberosum* L.) is the newest field crop undergoing conversion to a hybrid breeding system[1,2]. Following the success of other crops like sugar beet[3] and sunflower[4], both private and public ventures are endeavouring to convert this clonal tetraploid crop into a diploid chasing the same benefits that were originally observed in maize[5]. This conversion in potato is actually composed of two distinct and yet inter-related activities: (1) ploidy reduction from a tetraploid to diploid, and (2) the shift from outcrossing populations to structured inbred populations shaped around F1 hybrid performance. The benefits of the former primarily lie in the ease and speed of positive gene fixation in a disomic inheritance model over tetrasomic inheritance[6] with the latter comprising both technical (trait fixation in inbreds, exploitation of heterotic vigour) and commercial (intellectual property protection, consistent product quality and uniformity) related advantages[7]. Despite the allure of these propositions, it is only within the past decade that commercialization efforts of hybrid potato programs have come to pass due in part to overcoming genetic barriers like severe inbreeding depression, fertility anomalies, and ploidy barriers for secondary and tertiary gene pools[8–10].

It is well understood that while hybrid breeding systems bring many potential benefits to crop improvement strategies, this is not without significant cost. Hybrid breeding is known to suffer from longer breeding cycles in contrast to line breeding and necessitates specific populations for heterotic pool development[11,12]. A large aspect of that complexity lies in that the target for improvement is F1 performance and not inbred family *per se* yields due to poor correlations between the two. In the pre-genomics era, testcross evaluation was the crux of line advancement and population enrichment as it permitted the estimation of general and specific combining abilities[13]. It was not until the advent of molecular technologies that it became possible to develop inexpensive proxies in the place of field evaluation, first in the form of molecular markers for marker-assisted selection (MAS) not long followed by genome-wide regression using hundreds to thousands of markers for genomic selection (GS)[14,15]. These methods alone have been responsible for incredible gains for quantitative trait improvement in field and vegetable crops alike over the past 30 years.

While genome-wide selection is a powerful tool in quantitative trait improvement, it is not necessarily the most *efficient*. Genome-wide selection has been shown to work very well in both tetraploid and diploid potato populations[16,17], but the cost associated with training set development, dense molecular marker panels, and computational capacity make it a resource-intensive process. An interesting alternative would be identifying quantitative traits which could be selected solely on the basis of QTL information. The literature is ripe with examples of large-effect QTL responsible for controlling traits such as tuber shape[18], starch content[19], maturity[20,21] and even volume and weight of tubers[22,23]. Making these identified QTL the basis of selection in a MAS program could be a simple and cost-effective method for improving quantitative trait targets in potato. However, while the success of MAS is well understood in other crops and simulation[24,25], its utility in hybrid potato is currently unknown. We examine the following scenarios: a marker-assisted breeding

scenario which conducts selection based on identified QTL in the hybrid (`MA`), a genome-wide selection method (`GW`), and lastly a method which bases selection on 1 random SNP per chromosome (`PC`). These three strategies were assessed on their prediction accuracy for the tuber phenotype as well as on the relative efficiency with respect to genome-wide selection. With these results, we hope to identify if marker-assisted selection is a useful selection method in practice for complex traits in potato.

4.2 Material and methods

Hybrid Phenotypic Data

This study used a population of 785 genotyped F1 hybrids which had undergone field evaluation between 2019 and 2020. As described in[16], this dataset comprises four field trials in two trial locations (Heelsum and Est, Netherlands). Best linear unbiased predictions (BLUPs) for hybrids were computed using the following model:

$$y_{bfht} = \beta_f + \gamma_{fb} + a_{ht} + \varepsilon_{bfht} \tag{4.1}$$

where y is some component of yield for hybrid h within block b in trial location t and specific field trial f. y is modeled on a fixed trial (β_f) and block within-trial (γ_{fb}) effects, a random hybrid by trial (a_{ht}), and an independent and identically distributed (IID) residual term (ε_{bfht}). The hybrid by trial effect was modeled as Gaussian with an unstructured variance structure:

$$a \sim \mathrm{N}\left(\mathbf{0}, \Sigma_a\right) \qquad \Sigma_a = \begin{bmatrix} \sigma^2_{\mathrm{G(Est)}} & \sigma_{\mathrm{G(Est),G(Hee)}} \\ \sigma_{\mathrm{G(Hee),G(Est)}} & \sigma^2_{\mathrm{G(Hee)}} \end{bmatrix}$$

This model structure was chosen over simpler multi-environment models due to large observed differences in genetic variances during exploratory analysis. Such large discrepancies in variance were not observed for the residual term across trials so this retained its IID variance structure. Several variates were analyzed including total tuber yield (TY; $\mathrm{M}g \cdot \mathrm{Ha}^{-1}$), tuber dry matter content (DM; per cent), and average tuber volume (TV; cm^3).

Marker processing

Using the same dataset described in[16], marker data was collected on the 416 parental lines used in the creation of the evaluated hybrids. These lines were genotyped with a targeted genotyping-by-sequencing (GBS) technology marketed as SeqSNP and developed by LGC[26]. For this analysis, single nucleotide polymorphism's (SNPs) were extracted and filtered from the probeset and filtered based upon SNP missingness, minor allele frequency, and SNP density. These filter criteria left 1,040 remaining SNP's which were used to infer F1 hybrid genotypes for QTL mapping and genome-wide prediction.

QTL Scanning and Incorporation

QTL scans were conducted on all phenotypes using the hybrid BLUPs averaged over location and standard errors produced from ASReml-R's `predict()` interface along with SNP data. This was done using a forward-selection approach. The null model has the following form:

$$\hat{\mathbf{y}} = \mathbf{1}\mu + \mathrm{X_Q}\beta_Q + \mathrm{Z_g}u + e \tag{4.2}$$

where $\hat{\mathbf{y}}$ is the vector of hybrid predictions, μ is the global mean, β_Q is a vector of QTL fixed effects, u is a vector of random background genetic effects ($u \sim \mathrm{N}(0, \sigma_g^2)$), and e is a vector of residuals whose variance is structured from the first-stage model standard errors ($e \sim \mathrm{N}(0, \mathrm{R} = \mathrm{Diag}(\hat{\mathrm{se}}_1^2, \hat{\mathrm{se}}_2^2, \ldots, \hat{\mathrm{se}}_\mathrm{n}^2))$). During the first QTL scan, X_Q, the design matrix which contains SNP dosage of previously identified QTL, and β_Q will be empty. The design matrix for the background genetic effects, Z_g, is a diagonal matrix.

This model is tested against a scanning model:

$$\hat{\mathbf{y}} = \mathbf{1}\mu + \mathrm{X_Q}\beta_Q + \mathrm{Z_g}u + \mathrm{Z_{snp}}v + e \tag{4.3}$$

which differs only in the addition of one random term, v, the SNP effect ($v \sim \mathrm{N}(0, \sigma_{qtl}^2)$) being tested at any given scan. The forward-selection procedure consisted of: (1) fitting the null model in (Equation 4.2), (2) fitting the scanning model (Equation 4.3), (3) conducting a likelihood ratio test to assess the significance of v, (4) repeating steps 2-3 for all remaining SNPs, (5) selecting the SNP with the smallest p-value and placing it in X if it was smaller than the cutoff (defined by a Bonferroni threshold) and then repeating. If there were no remaining SNPs with sufficient evidence of putative QTL, the scan was concluded. A Wald test was then conducted on the final model to confirm the significance of all QTL in $\mathrm{X_Q}$; the effect size of all QTLs were extracted from this model (Figure 4.3). While some QTLs clustered within close proximity to one another, all QTL distances were above 1 Mb and thus considered independent of one another. Lastly, each QTL was fit in a single QTL model ($\hat{\mathbf{y}} = \mathbf{1}\mu + \mathrm{Z_g}u + \mathrm{Z_{snp}}v + e$) to estimate the proportion of QTL variance (Figure 4.2).

The number of QTL taken up in each model was recorded (Table 4.2) along with their significance (i.e. $-\log_{10}(p)$) during initial model scanning (Figure 4.1). All QTL modeling and phenotypic analyses were done using ASReml-R 4.2[27].

Genomic Prediction Models

Because one of the major research questions was related to testing the utility of a multi-QTL model in a marker-assisted selection program, we constructed multiple prediction models associated with distinct modeling scenarios (See Table 4.1). The first was a marker-assisted selection model (`MA`) which bases prediction upon the previously identified QTL (Table 4.2). All QTL which were found to be jointly significant were incorporated into

a traditional linear regression model to predict a candidate's value among our selected quantitative traits. This model was:

$$\hat{\mathbf{y}} = \mathbf{1}\mu + \mathrm{X}_Q\beta_Q + \varepsilon \tag{4.4}$$

where the hybrid BLUPs were modeled only on the global mean, and estimated QTL effects (β_Q) in each hybrid. This scenario was contrasted against a genome-wide based regression using 1000 SNPs (`GW`) and a model which picks one random SNP per chromosome (`PC`). The latter is meant to perform as a negative control to evaluate the minimum performance threshold from a SNPs linkage within a linkage group. Both of these models were trained using ridge regression and could be expressed as:

$$\hat{\mathbf{y}} = \mathbf{1}\mu + \mathrm{W}u + \varepsilon \tag{4.5}$$

where the SNP dosages are contained in matrix, W, and all marker effects in u. To assess prediction performance, all models were tested using 5-fold cross-validation. Cross-validation was replicated 100 times for each trait for the `MA` model and Pearson correlation coefficients were estimated for each fold and then averaged over each replicate. For models which introduced marker sampling variation (i.e., `GW` and `PC`), the replication was stratified over marker sampling ($n = 20$) and fold partitioning of hybrids ($m = 50$) for a total of 1000 replications. Similarly, the predictive performance was assessed via the Pearson correlation coefficient between the true and predicted hybrid performance and averaged over each fold. The average Pearson correlation coefficient were then compared across all models and traits (Figure 5.4). All prediction models were trained and tested using the `mixed.solve()` function provided through the rrBLUP R library[28].

Table 4.1: Breeding scenarios with descriptions of model parameters and cost associated with each.

Scenario code	Scenario	$\hat{\beta}$	$\hat{u}$	Cost
MA	Marker-assisted selection with identified QTLs	Identified QTL		\$0.15 per SNP per sample
GW	Genome-wide regression using mid-sized probe set	global mean	1000 genome-wide SNP's	\$150.00 per sample
PC	Selection with one random SNP on each chromosome	global mean	1 SNP per chromosome	\$1.80 per sample

Putting the model performance in the context of a breeding program, we examine `MA`, `PC`, and `GW` as alternative modeling methods for predicting the genetic value of 1000 novel F4 candidates for creating superior F1 candidates. This is done on the basis of marker effect estimation at the F1 hybrid level which would make inbred selection based on that line's genomically estimated breeding value. These schemas were evaluated based on the

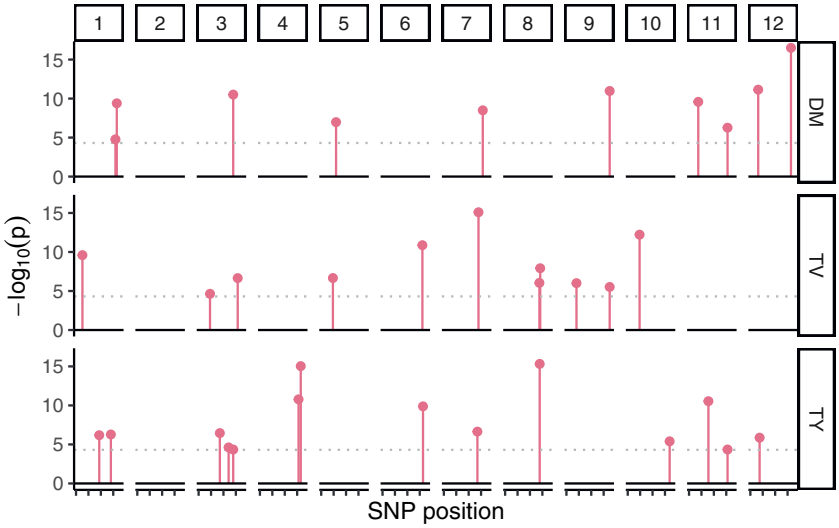


Figure 4.1: The physical positions and log_{10} probabilities found for all 35 QTL across potato genome for dry matter content (DM), tuber volume (TV), and total tuber yield (TY).

relative efficiency for the `MA` and `PC` scenarios relative to `GW` (Table 4.1). This is the ratio of the predicted selection response ($i\sigma_g r$) of each scenario over the genome-wide regression scenario[29]. Because the only difference between these scenarios is in the selection accuracy (r), the relative efficiency is then just the ratio of the selection accuracy from each scenario. This was extended further through the inclusion of genotyping costs of the 1000 candidates for each strategy assuming a SNP assay genotyping method. This was done by dividing the predicted selection response of each strategy by the log of the genotyping cost. While costs vary according to locale, number of samples, and specific technology, we chose pricing that appeared reasonable given trends over the past decade[30]. Therefore, each scenario was assessed with respect to the relative efficiency and the relative efficiency divided by the log of the genotyping costs (Table 4.3).

4.3 Results

QTL Scan

All scanning models converged successfully and picked up multiple QTL for each phenotype (Table 4.2). The final QTL models exhibited a better goodness of fit in contrast to the null model. This was reflected both in the change of the AIC as well as reduction in background genetic variance in each final model, though, this was trait dependent. Between 53 and 56% of the genetic variance was partitioned away from the background genetic term between the null and final model across traits. Among the traits tested, dry matter content appeared to improve most through the incorporation of the 10 identified QTL (Table 4.2).

In total, 35 QTL peaks were detected but given overlap between traits, this comprised 33 unique QTL (Table 4.2). Total tuber yield detected the largest number of QTLs (14 QTL) with tuber volume and dry matter content detecting 11 and 10, respectively. The QTL variance varied between 10 and 62% of the total genetic variance among all phenotypes but the median proportion of variance was highest in drymatter content (36.0%) followed by total tuber yield (30.1%) and average tuber volume (29%) (Figure 4.2). There were multiple QTL clusters detected for all three phenotypes comprising chromosome 1 for dry matter content (chr01:83523334 and chr01:86604310), chromosome 8 in tuber volume (chr08:46811267 and chr08:47889222), chromosome 4 in total yield (chr04:60035708 and chr04:64373508).

Table 4.2: Model parameters for the null and final QTL models for dry matter content (DM), average tuber volume (TV) and total tuber yield (TY). The Akaike Information Criterion (AIC), the change in AIC (ΔAIC), background genetic variance (σ_g^2), and per cent change in background genetic variance ($\Delta\%$) for the initial null and full QTL models. This is presented with the estimated global mean of each trait and the total number of QTL incorporated into the final model.

	AIC			σ_g^2				
Phenotype	Null model	Final model	Δ	Null model	Final model	$\Delta\%$	μ	Number of QTL
DM	1690.0	1299.1	390.9	2.5	1.2	52.9	20.7	10
TV	3650.0	3115.7	534.3	35.7	15.6	56.4	29.9	11
TY	3394.1	2913.8	480.4	24.6	11.1	54.6	14.5	14

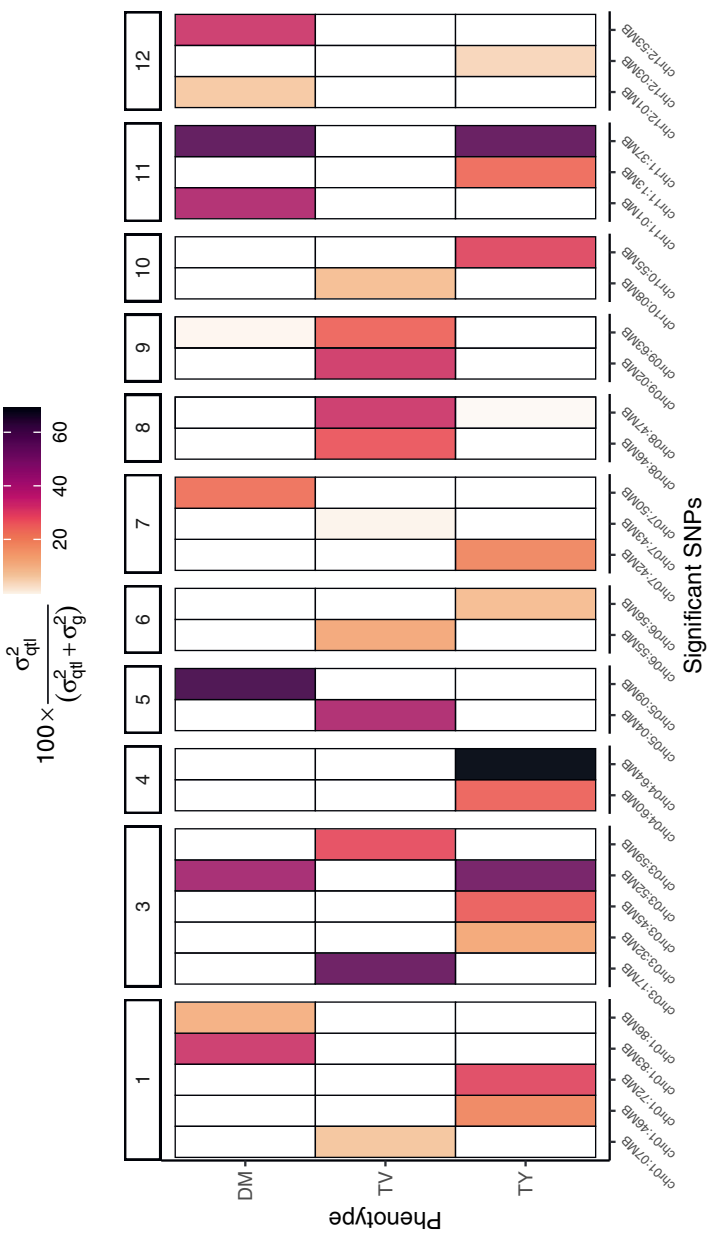


Figure 4.2: Proportion of captured genetic variance from each QTL in the scanning model among all 35 identified QTL separated per chromosome in dry matter content (DM), average tuber volume size (TV), and total tuber yield (TY).

Examining all detected QTL, varying QTL effect size was observed with respect to magnitude and direction (Figure 4.3). The largest QTL effects were observed in total tuber yield with effects as large as 30% of the mean. The QTL observed in dry matter content also tended to be smaller in magnitude than those observed in tuber volume and total tuber yield. Two correlated QTL were observed for dry matter content and total tuber yield on chromosome 3 (chr03:52583805) and on chromosome 11 (chr11:37618884) (Figure 4.1). Both QTLs had a relatively large QTL variance (Figure 4.2) and were negatively correlated between total tuber yield and dry matter content (Figure 4.3).

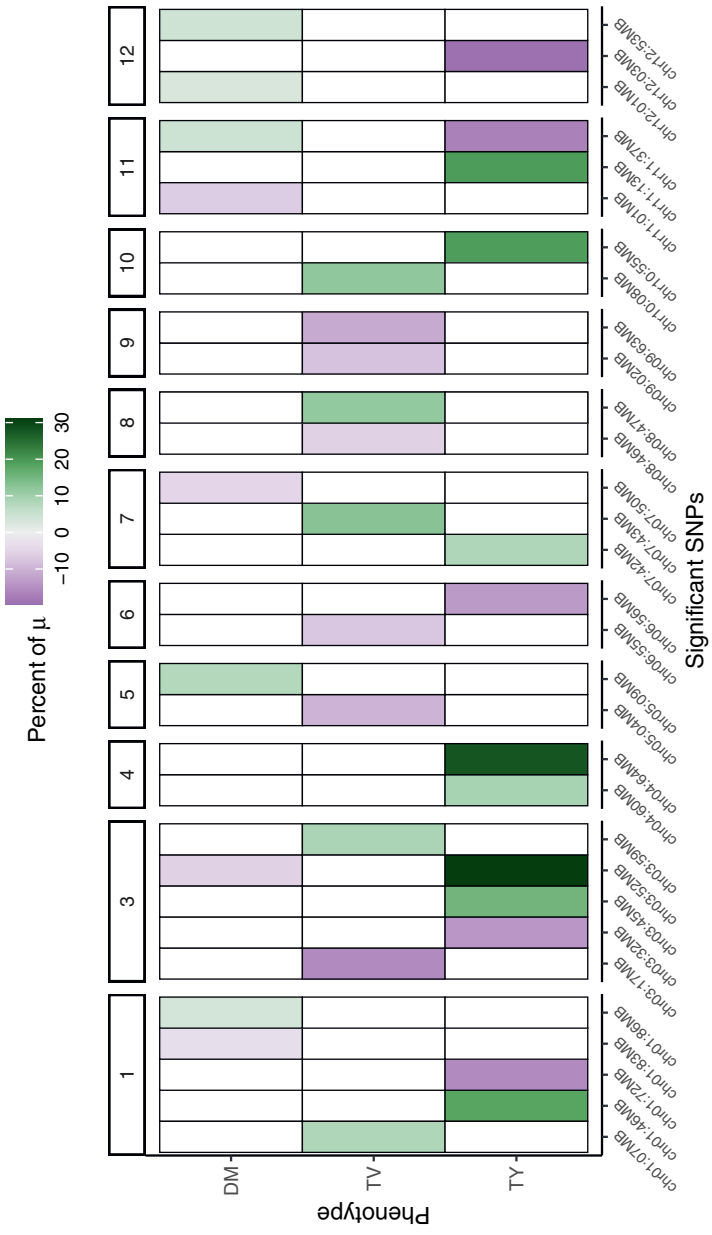


Figure 4.3: Marker effect size in β_{qtl} from the final QTL model among all 35 identified QTL separated per chromosome in dry matter content (DM), average tuber volume size (TV), and total tuber yield (TY).

Genomic Prediction Models

Turning to each predictive strategy, the GW scenario performed consistently better than all other strategies and represented the maximum prediction potential that could be attained (Figure 5.4). It also appears to closely mirror the phenotype heritabilities and performance observed in previous studies (See Adams, De Vries, and Van Eeuwijk [16]). Conversely, the model used in PC were consistently the worst performing with observed prediction accuracies between 0.24 and 0.34 across phenotypes. The MA models not only surpassed the negative control (PC), but followed closely with the performance observed in the GW model with ρ's between 0.53 and 0.68 across traits. The PC strategy had the highest variability in ρ during cross-validation due likely to the large probe sampling variance. Examining these results relative to the GW scenario, MA had a far larger relative efficiency for the three traits targets (0.89 - 0.95) in contrast to PC (0.42 - 0.43) (Table 4.3). Looking to genotyping costs, the cost ratios varied very little between MA and PC strategies; both strategies were within two orders of magnitude cheaper than the GW genotyping strategy. All relative efficiencies were less than 1 but when genotyping costs were considered, the MA was considerably higher for all traits (1.41 - 1.48) indicative of higher selection efficiencies per log cost. This was in contrast to the PC scenario which had very low relative efficiency per log cost.

4.4 Discussion

An important exercise for any breeding program is the careful analysis of goals for trait improvement in light of the genetic variation present, the statistical architecture of the trait, and available resources available to a program. Examining the particular question of genotyping technology and selection methods, we provide a template for evaluating the efficiency of breeding strategies. In this paper, we were able to detect multiple QTL which captured a large portion of the genetic variance in multiple tuber phenotypes. The merit of these QTL were then tested through against a genome-wide regression strategy and a random marker-based selection strategy as a negative control. Selection responses were examined for all scenarios with respect to unit gain and unit gain per unit cost. This represents an invaluable procedure for assessing the benefits of two selection methods in an operational breeding program.

Selecting Quantitative traits

The first criteria for the success of quantitative trait selection is the degree of genetic variability for the trait in question, in other words, the trait's heritability. Based upon the results of this study and previous works[31] we know that all phenotypes studied here are highly heritable and capable of genetic improvement. Based on prediction accuracy (Figure 5.4) and consideration of selection costs (Table 4.3), we found credible evidence that both MAS and genome-wide selection would be successful selection methods for the traits studied here. Regarding genome-wide selection, this was in line with what other studies have observed for these traits in potato[16,17]. However, the efficiency of the MA strategy for quantitative trait selection needs further consideration. While MAS

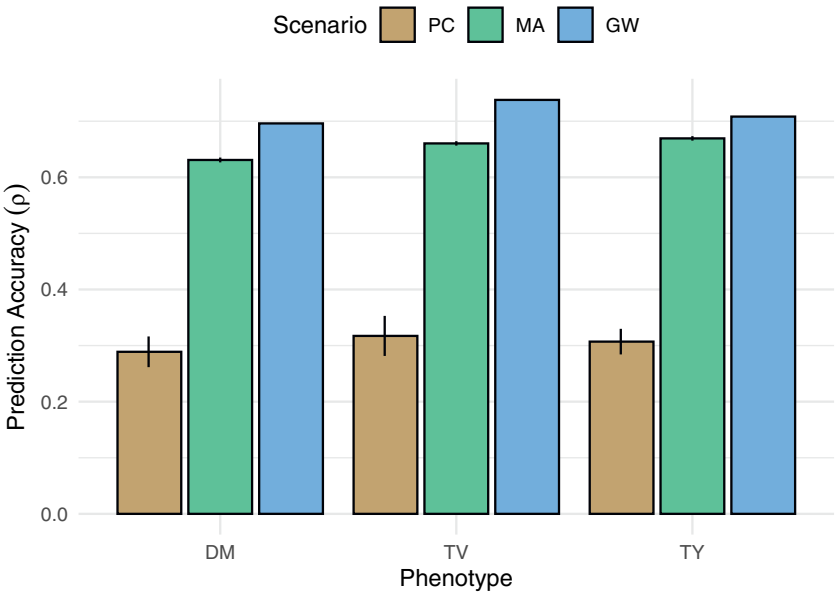


Figure 4.4: Prediction accuracy results from 5-fold cross validation for qtl mapping (qtl), genome-wide regression (GW), and marker-assisted selection with 12 random SNPs (PC). Error bars contain one standard deviation around the average Pearson correlation coefficient.

is an older technology and prone to lower selection accuracies, it is still potentially an effective selection method if sufficient genetic variation can be found. In our QTL mapping procedure, over half the genetic variation was accounted for in the final QTL model, and yet, the relative efficiency was as high as 0.95. This could be a consequence of incorporating large effect QTL which carry most of the relevant information for prediction. This was in contrast to the `PC` control which, while having a similar number of SNPs used for selection, offered little in way of selection response. So given that there are still large effect QTL segregating among candidates, the `MA` strategy is suitable in selection if well designed markers are being used.

Table 4.3: The relative efficiency (RE), genotyping cost ratio (CR) and relative efficiency per log cost (RELC) for the marker-assisted selection (MA) and the single marker per chromosome strategies (PC) relative to the genome-wide selection scenario (GW) for dry matter content (DM), average tuber volume (TV), and total tuber yield (TY).

	MA			PC		
Phenotype	RE	CR	RELC	RE	CR	RELC
DM	0.91	0.010	1.48	0.42	0.012	0.66
TV	0.89	0.013	1.41	0.43	0.012	0.68
TY	0.95	0.014	1.47	0.43	0.012	0.69

MAS versus genome-wide selection: Costs and complexity

When considering which of these strategies would be more appropriate for employing in a hybrid potato breeding program, multiple questions need to be considered. Most of these center around the maturity of the breeding program and the priority of targets. Most crop breeding programs are more conservative with genetic variation loss, and therefore, would likely utilize a selection strategy which guarantees the highest selection accuracy possible[32,33]. In the context of these two tested scenarios (`MA` and `GW`), this would lean towards the genome-wide prediction approach. However, there are other programs that could be less risk-averse with regard to loss of genetic variation so long as it allows for selection response (or unit selection response per unit cost) to be maximized by some other means. We observed that when considering genotyping costs, the `MA` strategy was more efficient per log cost for all three trait targets (Table 4.3). When a cost penalty is introduced, any affordable technology that borders with the performance of genome-wide regression becomes immediately interesting for adoption. So for breeding programs which value greater or more cost-efficient short-term gains, a marker-assisted selection approach is worth consideration for quantitative trait selection.

Operational considerations

These scenarios are presented as is with no formal consideration of other operational questions which could be relevant for their effective adoption. For example, we used a population where QTLs could be identified, however, this is of course not guaranteed. The success of QTL identification in a breeding program is dependent on many interrelated questions: population mapping or family mapping[34], having dedicated mapping populations or using extant material, and the number of individuals used[14]. Even after QTL identification, there are common pitfalls in maintaining a marker in a MAS program such as recombination between the QTL and marker, ensuring universal probe annealing across relevant genetic backgrounds, and maintaining these targets alongside other important genes necessary for a candidate's portfolio. Genome-wide prediction methods in this regard, are more straightforward, as model training, testing, validation, and retraining are regular procedures as breeding populations develop over time. However, an unappreciated logistical challenge with such predictive workflows is proper investment in data management and well engineered data pipelines as the size of datasets accumulate each year. The creation and maintenance of such digital infrastructure for genomic workloads requires expertise outside of the genetical and statistical domain and could be a liability if not sufficiently planned accordingly[35,36].

4.5 Conclusions

Complex traits in field crops include some of the most economically valuable targets for trait improvement such as biomass partitioning, abiotic stress resistance, and crop yield. Therefore, identifying schema which maximize the rate of genetic progress in candidate selection is an indispensable activity for breeding programs. Here we focused our attention on three important yield phenotypes in potato and found that MAS could serve as an

affordable selection methodology in the place of genome-wide predictive methods. Such a low-cost technology lends itself to allocation bottlenecks common to breeding programs where the number of candidates are large, a scenario common to multiple stages of a hybrid breeding program. Tools like MAS are thus still a valuable part of the breeders toolbox for genetic improvement in hybrid potato.

Bibliography

1. De Vries ME, Adams JR, Eggers Ej, Ying S, Stockem JE, Kacheyo OC, Van Dijk LCM, Khera P, Bachem CW, Lindhout P, and Van Der Vossen EAG. Converting Hybrid Potato Breeding Science into Practice. Plants. 2023 Jan; 12:230. DOI: 10.3390/plants12020230. Available from: https://www.mdpi.com/2223-7747/12/2/230 [Accessed on: 2024 Jun 3]
2. Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS, Bamberg JB, Thompson AL, Bizimungu B, Holm DG, Brown CR, Haynes KG, Sathuvalli VR, Veilleux RE, Creighton Miller J, Bradeen JM, and Jiang J. Reinventing Potato as a Diploid Inbred Line-Based Crop. Crop Science. 2016 Jul 1; 56:1412–22. DOI: 10.2135/cropsci2015.12.0740
3. McGrath JM and Panella L. Sugar Beet Breeding. *Plant Breeding Reviews*. Ed. by Goldman I. 1st ed. Wiley, 2018 Dec :167–218. DOI: 10.1002/9781119521358.ch5. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9781119521358.ch5 [Accessed on: 2024 Mar 20]
4. Dimitrijevic A and Horn R. Sunflower Hybrid Breeding: From Markers to Genomic Selection. Frontiers in Plant Science. 2018 Jan; 8:2238. DOI: 10.3389/fpls.2017.02238. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2017.02238/full [Accessed on: 2024 Mar 20]
5. Duvick D. Heterosis: Feeding People and Protecting Natural Resources. *The Genetics and Exploitation of Heterosis in Crops*. 1997 :6–7
6. Hougas R and Peloquin S. The Potential of Potato Haploids in Breeding and Genetic Research. American Potato Journal. 1958; 35:701–7
7. Wright H. Commercial Hybrid Seed Production. *Hybridization of Crop Plants*. Ed. by Fehr WR and Hadley HH. Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, 1980 Jan :161–76. DOI: 10.2135/1980.hybridizationofcrops.c8. Available from: http://doi.wiley.com/10.2135/1980.hybridizationofcrops.c8 [Accessed on: 2024 May 7]
8. Zhang C, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, and Huang S. The Genetic Basis of Inbreeding Depression in Potato. Nature Genetics. 2019 Mar 14; 51:374–8. DOI: 10.1038/s41588-018-0319-1. Available from: http://www.nature.com/articles/s41588-018-0319-1 [Accessed on: 2019 Mar 20]
9. Endelman JB and Jansky SH. Genetic Mapping with an Inbred Line-Derived F2 Population in Potato. Theoretical and Applied Genetics. 2016 May; 129:935–43.

DOI: 10.1007/s00122-016-2673-7. Available from: http://link.springer.com/10.1007/s00122-016-2673-7 [Accessed on: 2024 May 7]

10. Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, and Huang S. Genome Design of Hybrid Potato. Cell. 2021 Jul 22; 184:3873–3883.e12. DOI: 10.1016/J.CELL.2021.06.006
11. Labroo MR, Studer AJ, and Rutkoski JE. Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. Frontiers in Genetics. 2021 Feb 24; 12:234. DOI: 10.3389/FGENE.2021.643761
12. Longin CFH, Gowda M, Mühleisen J, Ebmeyer E, Kazman E, Schachschneider R, Schacht J, Kirchhoff M, Zhao Y, and Reif JC. Hybrid Wheat: Quantitative Genetic Parameters and Consequences for the Design of Breeding Programs. Theoretical and Applied Genetics. 2013 Nov; 126:2791–801. DOI: 10.1007/s00122-013-2172-z. Available from: http://link.springer.com/10.1007/s00122-013-2172-z [Accessed on: 2024 May 6]
13. Carangal VR, Ali SM, Koble AF, Rinke EH, and Sentz JC. Comparison of S 1 with Testcross Evaluation for Recurrent Selection in Maize 1. Crop Science. 1971 Sep; 11:658–61. DOI: 10.2135/cropsci1971.0011183X001100050016x. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci1971.0011183X001100050016x [Accessed on: 2024 May 6]
14. Lande R and Thompson R. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. Genetics. 1990 Mar; 124:743–56. DOI: 10.1093/genetics/124.3.743. Available from: https://academic.oup.com/genetics/article/124/3/743/5999957 [Accessed on: 2024 Mar 15]
15. Meuwissen THE, Hayes BJ, and Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics. 2001 Apr 1; 157:1819–29. DOI: 10.1093/genetics/157.4.1819. Available from: https://academic.oup.com/genetics/article/157/4/1819/6048353 [Accessed on: 2022 Nov 13]
16. Adams J, De Vries M, and Van Eeuwijk F. Efficient Genomic Prediction of Yield and Dry Matter in Hybrid Potato. Plants. 2023 Jul 11; 12:2617. DOI: 10.3390/plants12142617. Available from: https://www.mdpi.com/2223-7747/12/14/2617
17. Wilson S, Malosetti M, Maliepaard C, Mulder HA, Visser RGF, and van Eeuwijk F. Training Set Construction for Genomic Prediction in Auto-Tetraploids: An Example in Potato. Frontiers in Plant Science. 2021; 12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.771075 [Accessed on: 2022 Nov 15]
18. Eck HJV, Oortwijn ME, Terpstra IR, Lieshout NHV, Knaap EVD, Willemsen JH, and Bachem CW. Engineering of Tuber Shape in Potato (Solanum Tuberosum) with Marker Assisted Breeding or Genetic Modification Using StOFP20. preprint. In Review, 2022 Jul. DOI: 10.21203/rs.3.rs-1807189/v1. Available from: https://www.researchsquare.com/article/rs-1807189/v1 [Accessed on: 2024 Jan 4]
19. Li H, Jiang S, Li C, Liu L, Lin Z, He H, Deng XW, Zhang Z, and Wang X. The Hybrid Protein Interactome Contributes to Rice Heterosis as Epistatic Effects. The Plant Journal. 2019 Dec :tpj.14616. DOI: 10.1111/tpj.14616. Available from:

https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.14616 [Accessed on: 2019 Dec 29]

20. Li X, Xu J, Duan S, Zhang J, Bian C, Hu J, Li G, and Jin L. Mapping and QTL Analysis of Early-Maturity Traits in Tetraploid Potato (Solanum Tuberosum L.) International Journal of Molecular Sciences. 2018 Oct; 19:3065. DOI: 10.3390/ijms19103065. Available from: http://www.mdpi.com/1422-0067/19/10/3065 [Accessed on: 2024 Mar 20]
21. Massa AN, Manrique-Carpintero NC, Coombs JJ, Zarka DG, Boone AE, Kirk WW, Hackett CA, Bryan GJ, and Douches DS. Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (*Solanum* Tuberosum L.) G3 Genes\textbarGenomes\textbarGenetics. 2015 Nov; 5:2357–64. DOI: 10.1534/g3.115.019646. Available from: https://academic.oup.com/g3journal/article/5/11/2357/6025373 [Accessed on: 2024 Mar 20]
22. Korontzis G, Malosetti M, Zheng C, Maliepaard C, Mulder HA, Lindhout P, Veerkamp RF, and van Eeuwijk FA. QTL Detection in a Pedigreed Breeding Population of Diploid Potato. Euphytica. 2020 Aug; 216:145. DOI: 10.1007/s10681-020-02674-y. Available from: https://doi.org/10.1007/s10681-020-02674-y [Accessed on: 2023 Jan 3]
23. Clot CR, Wang X, Koopman J, Navarro AT, Bucher J, Visser RGF, Finkers R, and Van Eck HJ. High-Density Linkage Map Constructed from a Skim Sequenced Diploid Potato Population Reveals Transmission Distortion and QTLs for Tuber Yield and Pollen Shed. Potato Research. 2024 Mar; 67:139–63. DOI: 10.1007/s11540-023-09627-7. Available from: https://link.springer.com/10.1007/s11540-023-09627-7 [Accessed on: 2024 May 7]
24. Podlich DW, Winkler CR, and Cooper M. Mapping As You Go: An Effective Approach for Marker-Assisted Selection of Complex Traits. Crop Science. 2004 Sep; 44:1560–71. DOI: 10.2135/cropsci2004.1560. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci2004.1560 [Accessed on: 2024 Mar 20]
25. Van Berloo R, Van Heusden S, Bovy A, Meijer-Dekens F, Lindhout P, and Van Eeuwijk F. Genetic Research in a Public–Private Research Consortium: Prospects for Indirect Use of Elite Breeding Germplasm in Academic Research. Euphytica. 2008 May; 161:293–300. DOI: 10.1007/s10681-007-9519-y. Available from: https://link.springer.com/10.1007/s10681-007-9519-y [Accessed on: 2024 Jul 2]
26. LGC. SeqSNP Targeted GBS as Alternative for Array Genotyping in Routine Breeding Programs [White Paper]. Biosearch Technologies, 2019. Available from: https://biosearch-cdn.azureedge.net/assetsv6/seqsnp-tgbs-alternative-genotyping-routine-breeding-programs.pdf
27. Butler DG, Cullis BR, Gilmour AR, Gogel BJ, and Thompson R. ASReml-R Reference Manual Version 4. Hemel Hempstead, HP1 1ES: VSN International Ltd, 2017. Available from: http://www.homepages.ed.ac.uk/iwhite/asreml/uop. [Accessed on: 2021 Jan 19]

28. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome. 2011; 4. DOI: 10.3835/plantgenome2011.08.0024. Available from: https://onlinelibrary.wiley.com/doi/abs/10.3835/plantgenome2011.08.0024 [Accessed on: 2022 Dec 9]
29. Krchov LM and Bernardo R. Relative Efficiency of Genomewide Selection for Testcross Performance of Doubled Haploid Lines in a Maize Breeding Program. Crop Science. 2015 Sep; 55:2091–9. DOI: 10.2135/cropsci2015.01.0064. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci2015.01.0064 [Accessed on: 2024 Apr 10]
30. Semagn K, Babu R, Hearne S, and Olsen M. Single Nucleotide Polymorphism Genotyping Using Kompetitive Allele Specific PCR (KASP): Overview of the Technology and Its Application in Crop Improvement. Molecular Breeding. 2014 Jan; 33:1–14. DOI: 10.1007/s11032-013-9917-x. Available from: http://link.springer.com/10.1007/s11032-013-9917-x [Accessed on: 2024 Apr 10]
31. Adams JR, de Vries ME, Zheng C, and van Eeuwijk FA. Little Heterosis Found in Diploid Hybrid Potato: The Genetic Underpinnings of a New Hybrid Crop. G3: Genes, Genomes, Genetics. 2022 Jun; 12. DOI: 10.1093/g3journal/jkac076. PMID: 35460241
32. Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, and Glenn KC. Genetic Diversity Is Indispensable for Plant Breeding to Improve Crops. Crop Science. 2021 Mar; 61:839–52. DOI: 10.1002/csc2.20377. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.20377 [Accessed on: 2024 Apr 28]
33. Rasmusson DC and Phillips RL. Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. Crop Science. 1997 Mar; 37:303–10. DOI: 10.2135/cropsci1997.0011183X003700020001x. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci1997.0011183X003700020001x [Accessed on: 2024 May 29]
34. Würschum T. Mapping QTL for Agronomic Traits in Breeding Populations. Theoretical and Applied Genetics. 2012 Jul; 125:201–10. DOI: 10.1007/s00122-012-1887-6. Available from: http://link.springer.com/10.1007/s00122-012-1887-6 [Accessed on: 2024 Jan 6]
35. O'Driscoll A, Daugelaite J, and Sleator RD. 'Big Data', Hadoop and Cloud Computing in Genomics. Journal of Biomedical Informatics. 2013 Oct; 46:774–81. DOI: 10.1016/j.jbi.2013.07.001. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1532046413001007 [Accessed on: 2024 Mar 15]
36. Mulder NJ, Adebiyi E, Adebiyi M, Adeyemi S, Ahmed A, Ahmed R, Akanle B, Alibi M, Armstrong DL, Aron S, Ashano E, Baichoo S, Benkahla A, Brown DK, Chimusa ER, Fadlelmola FM, Falola D, Fatumo S, Ghedira K, Ghouila A, Hazelhurst S, Isewon I, Jung S, Kassim SK, Kayondo JK, Mbiyavanga M, Meintjes A, Mohammed S, Mosaku A, Moussa A, Muhammd M, Mungloo-Dilmohamud Z, Nashiru O, Odia T, Okafor A, Oladipo O, Osamor V, Oyelade J, Sadki K, Salifu SP, Soyemi J, Panji S, Radouani F, Souiai O, Tastan Bishop Ö, and Members Of The HAfrica Consortium As THabC. Development of Bioinformatics Infrastructure for Genomics Research.

Global Heart. 2017 Jun; 12:91. DOI: 10.1016/j.gheart.2017.01.005. Available from: https://globalheartjournal.com/article/10.1016/j.gheart.2017.01.005/ [Accessed on: 2024 Mar 15]

5

Evaluating Multiallelism for Genomic Prediction Applications in Potato

Abstract

Potato breeding must be improved in order to hasten the release of higher-yielding and more robust potato varieties. The development and wider adoption of genetic tools like genomic prediction depend on high quality and high-throughput genotyping platforms in order to estimate the genetic value of a candidate. Rapid improvement of the potato genome sequence and genotyping platforms are both a boon and problem for breeders as a higher volume of high quality marker data needs proper methodological applications to make it useful. Of specific interest to the authors of this paper is the use of multiallelic marker information, whether identity-by-state (using haplotags, HT) or identity-by-descent (IBD), and if it offers benefits over that of biallelic variants (using single nucleotide polymorphisms, SNPs) for the prediction of genomically estimated breeding values for tuber yield in potato. Consideration over model choice is also worthwhile given the hierarchical structure of multiple alleles to a given locus must be interrogated to ensure a proper representation of the statistical architecture. We considered traditional shrinkage-based models (ridge regression, LASSO), group-wise modified models (grouped LASSO), and advanced kernel approaches (Gaussian Kernel). Using phenotypic and genotypic data from a hybrid potato breeding program, we found that multiallelic-based predictors could be used to achieve adequate prediction accuracy for both IBD (total tuber yield: 0.65 - 0.67) and haplotag (total tuber yield: 0.71 - 0.75) probesets with the latter being generally more accurate in total tuber yield, tuber size and dry matter content. All prediction models using the biallelic probeset (SNP) were the most performant across all traits. Model choice had a significant impact on prediction accuracy and error with the Transformed Gaussian Kernel performing best across each trait. These findings suggest that while multiallelic marker information can offer better discriminative power in many breeding related applications, this is not necessarily the case for model-based genetic assessment. To our knowledge, this research represents a first evaluation of multiallelic predictors for genomic prediction applications in potato.

5.1 Introduction

Genetic improvement in potato (*Solanum tuberosum*) is a topic which has garnered much interest in recent decades. It includes issues related to large progeny variances, choice of propagation systems, significant genotype by environment interactions, and the fundamental biology of the crop[1–3]. For over a century, the challenges around potato's evolution in particular have come to the forefront beginning with its notable aversion to inbreeding and tetrasomic inheritance[4,5]. It was not until the advent of technologies like molecular sequencing that the effects of potato's polyploidization were seen in full with strong evidence of accumulated deleterious mutations, and genomic variation more generally[6–9]. This is exemplified by a large number of SNPs/indels (1 SNP variant per 29 basepairs) as well as a large quantity of structural variants (over 560,000 in one pan-genome) when sampling the Solanum genepool[10,11]. This also extends to important functional regions in potato. Examples include genes controlling maturity (*stCDF*), tuber shape (*StOFP20*), and self-incompatibility (*sli*)[12–14].

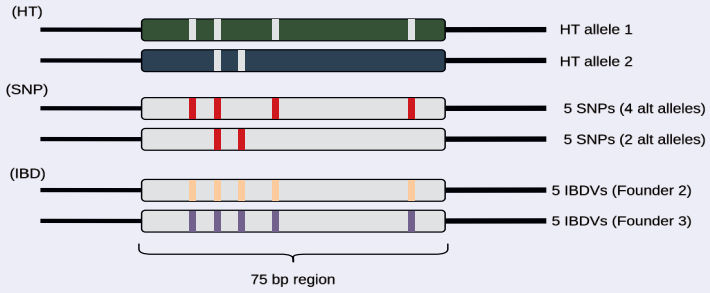
The state of functional variation in the Solanum genepool naturally has implications for genetic applications in potato. Since the release of the first potato genome, multiple molecular strategies have been deployed successfully including KASP, microarrays, untargeted genotyping-by-sequencing (GBS), and recently targeted GBS methods[15–18]. The latter approach is especially interesting for potato as it makes marker design amenable to hyper-variable regions where multiple haplotypes are present. This is already being leveraged in platforms like PotatoMASH (Potato Multi-Allele Scanning Haplotags), a targeted amplicon sequencing strategy which was developed with the express purpose of genetic discovery in potato[19]. The fundamental unit of capture are a group of variants that constitute a short-range haplotype, or to use the language of[20], a *haplotag*. The use of haplotags over traditional SNP variants has been shown to offer better statistical power in genetic modelling applications (e.g. GWAS) and have already been used for QTL discovery for multiple traits in potato including unreduced pollen, tuber eye depth, and tuber number[21–23].

If we want to consider multiallelic effects in potato, haplotags represent an identity-by-state (IBS) approach where any estimated effect is based upon the sampled DNA sequence. An alternative to this would be to infer the ancestral state of that piece of DNA relative to it's population history. Within population genetics, this is otherwise known as identity-by-descent (IBD)[24]. Increasing interest in IBD-based methods have coincided with the emergence of popular multi-parental population designs such as the NAM or MAGIC which enable parametrizing genetic models according to multiple founder effects[25,26]. Over the past decade, this interest has extended from structured balanced pedigrees to that of arbitrary design thanks to Hidden-Markov based approaches such as those found in RABBIT[27]. These methods become immediately interesting with deeper and more complex pedigrees such as those found in breeding programs[28,29]. Despite interest by breeders, there are scant studies which have evaluated their feasibility in a breeding program where strong forces like genetic drift and selection make breeding populations deviate significantly from these more idealized multi-parental populations[26,30].

Table 1 Description of four probesets used in hybrid prediction

(Top) The four predictor sets with the number of hybrid observations, the number of loci, and the number of predictors derived from each marker type. These include the original short-sequence based haplotags (HT), identity-by-state single nuclear polymorphisms (SNP), sparse single nuclear polymorphisms (sSNP), and the founder-derived identity-by-descent variants (IBD or IBDVs). (Bottom) Diagram of the haplotag model for all 811 probes used in this study. The amplicon sequence is composed of a 75 base pair target region which contains a variable number of distinguishing variants. The haplotag alleles are treated as is while both SNP and IBD datasets are based upon the constituent variants. The IBDVs are then the probability of a variant in question having descended from a given founder line. In this example, this region is heterozygous for founders 2 and 3 for all five variants within the amplicon sequence.

ID	Hybrids	Markers	Predictors
HT	783	811	2596
SNP	783	2612	2612
sSNP	783	1748	1748
IBD	783	1748	19518



It is clear that both haplotag and IBD variants have been useful in many genetic applications in potato[31,32]. However, to date, neither have been tested in genome-wide prediction for diploid or polyploid potato. There has also been little examination at benefits of IBD marker information over that of IBS when both leverage multiallelism. We therefore consider germplasm from a diploid potato hybrid breeding program to examine the following questions: (1) Whether multiallelic markers are more informative then their biallelic counterparts, (2) Are there statistical model paradigms which better exploit the multiallelic markers, and (3) whether IBD markers are more useful over IBS markers in a hybrid breeding schema. We interrogate these questions using a targeted GBS amplicon platform to derive haplotags (HT), SNP variants (SNP), estimated IBD profiles (IBD), and a SNP subset (sSNP) to evaluate benefits of a particular marker paradigm (See Table 1). We investigate popular shrinkage based methods (e.g. ridge regression, LASSO, Kernel regression) as well as group-wise models for the multiallelic markersets (group-LASSO).

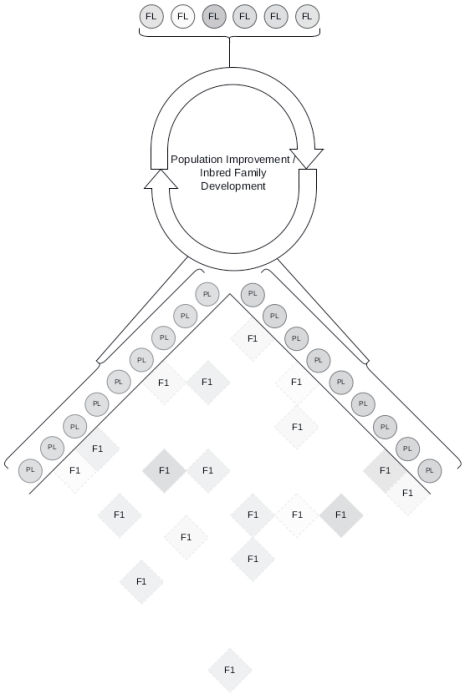


Figure 5.1: Diagram of a simple hybrid breeding schema with founder lines (FLs) at pedigree origin, descendent parent lines (PLs) at the crossing block and derived hybrid crosses (F1).

We measure the impact of marker and model choice in a hybrid breeding scenario where we predict F1 yield performance for several tuber traits in potato seedlings (Figure 5.1). Similar to previous studies[33], we restrict ourselves to three important yield components in potato for genetic interrogation, i.e., total tuber yield (TY), average tuber volume (TV), and tuber dry matter content (DM).

5.2 Material & Methods

Phenotypic Description

For this study, we utilized data from an active hybrid breeding program (Figure 5.1). We used 1,038 F1 hybrids created from intercrossing of 554 parent lines. These lines were developed through pedigree selection of an initial inter-crossing of founder lines followed by subsequent rounds of selfing, and strong selection pressure[34]. These hybrids were evaluated in four yield trials in 2019 and 2020, all of which were true potato seed (TPS) transplants. These four trials have been described in full in[33] with the same spatial modelling strategy leveraged here to derive the best linear unbiased predictions (BLUPs) for each hybrid over all trials. We chose three phenotypes which demonstrated consistent heritabilities in this population. This included tuber dry matter content (DM), average tuber volume (TV), and total tuber yield (TY). These BLUPs were then used in the

following phenotypic model:

$$\tilde{y} = \mu + g_i + g_j + s_{ij} + \varepsilon_{ij}$$

where $\tilde{y}$ is the hybrid BLUP, μ is the across-trial mean, g_i and g_j are the general combining abilities of the maternal and paternal lines i and j, respectively, s_{ij} is the specific combining ability of those parents, and ε_{ij} is the residual of that cross. The general and specific combining abilities were treated as random effects and considered IID ($g \sim \mathrm{N}(0, \sigma^2_{gca})$ $\quad s \sim \mathrm{N}(0, \sigma^2_{sca})$). The pooled residual variance across the four trials was used to estimate the residual variance (σ^2_{ε}). From this model, we extracted GCAs and variances for later reference (Figure 5.2).

Marker Analysis

Using the dataset described in[33], marker data was collected on 416 of the original 554 parental lines along with 13 founder lines which gave rise to this population. These lines were genotyped with a targeted genotyping-by-sequencing technology (SeqSNP) developed by LGC[35] which included 811 targeted probes with a varying number of haplotypes per haplotag locus (on average, 6 haplotypes per haplotag). All other marker predictors were derived from these haplotag sequences (HT). Variants were extracted from the HTs for the 416 inbred lines resulting in 2612 SNP predictors which became our IBS SNP probeset. Basic variant filtering was performed on the SNP set including minor allele frequency (maf ≥ 0.05) and missingness (miss ≤ 0.1) with mean imputation performed on missing calls.

These 2612 SNPs together with their pedigree were then used to estimate identity by descent (IBD) profiles with respect to the 13 genotyped founders. This was done using a Hidden Markov Modelling framework, RABBIT, a modular genetics software implemented in Julia with linkage map, marker quality control, simulation, and other genetic facilities[27]. We constructed a simple analytical pipeline composed from three function calls beginning first with `magicfilter` followed by `magicimpute`, and finally `magicreconstruct` using the forward-backward algorithm. At model convergence, we assessed the haplotype probabilities (i.e. The probability of a founder's DNA at a given locus), and removed regions of low confidence ($p < 0.5$). Lastly, we derived the expected founder allele dosage in the hybrid from their parental haplotype expectations (T) and the pedigree design matrix (Z) through simple matrix multiplication (ZT). In our analysis, we found evidence of increased sparsity in specific regions, particularly on chromosomes 1 and 5 (Supplementary Section 1.). This complicated any comparison of the IBD and SNP datasets as any difference in model performance could now be the product of ascertainment bias[36]. To ensure that our evaluation of IBD-based predictors was valid, we filtered on those SNP variants which passed through both RABBIT's quality control and post-pipeline filtering. This marker sparse SNP markerset (sSNP) was then tested allowing for a proper comparison to the IBD set.

Both SNP and sSNP predictors were encoded by their minor allele dosage {AA = 0, Aa = 1, aa = 2} with each site centered around zero by subtracting $2p_i$. However, for both

multiallelic datasets (HT and IBD) we used the additive multi-allelic encoding prescribed by[37] and[38] for the genetic partitioning of multi-allelic genotypes. First, an incidence matrix was constructed from the genotypes, or what is often referred to as *one-hot encoding* in the machine learning literature. Because we were only interested in estimating the $h-1$ allelic effects for each locus (otherwise the model would be over-specified), incidences of the most predominant allele were removed. We then took twice the allelic frequency of each haplotype at each locus and subtract the allelic dosages from it to center each predictor around zero (See Supplementary Section 2). Hybrid allele and haplotype dosages for SNP, sSNP, and HT markersets were constructed analogously to the IBD set with the parent allele dosages and pedigree design matrix. In the end, we were left with four different marker sets for testing (Table 1).

Statistical Modeling

To evaluate the general benefits of multiallelic information, we tested multiple modelling frameworks. All models here are extensions of the standard shrinkage-based methods where we regress our hybrid performance directly on some set of predictors ($\tilde{y} = \mathrm{X}\beta + e$). The only difference lies in how β was estimated and the contents of X. Beginning with the simplest, we conducted a ridge regression based analysis where the solution for β (i.e. our allele substitution effects) are found with:

$$_{\beta}||y - \mathrm{X}\beta||_2^2 + \lambda||\beta||_2^2$$

This is tantamount to ordinary least squares (OLS) with the addition of a shrinkage parameter (λ) with L_2 regularization on β. A popular alternative to ridge regression is the least absolute shrinkage and selection operator, or the LASSO:

$$_{\beta}\frac{1}{2}||y - \mathrm{X}\beta||_2^2 + \lambda||\beta||_1$$

It deviates from the ridge only in the extra ℓ_1 penalty on β which is equivalent to $\sum_i^p |\beta_i|$. This ensures that some predictors are selected while others are fixed to 0. For both ridge regression and LASSO models, we used the `glmnet` R package which invokes an elastic net approach where the ridge ($\alpha = 1$) and lasso ($\alpha = 0$) are special cases of a general bridge model[39].

Because our IBD and HT predictors have an inherent biological structure between them, it would be worthwhile to have it reflected in the model as well. If there is a particular region or gene which has some effect on a phenotype, it would be useful to group all alleles for that locus. One effective approach for this would be the group Lasso[40]. Adapting the same notation from above, the group LASSO can be expressed as:

$$_{\beta}\frac{1}{2}||y - \mathrm{X}\beta||_2^2 + \lambda\sum_{j=1}^{J}\sqrt{p_j}||\beta_j||_2$$

where the ℓ_1 penalty has been replaced with an ℓ_2 penalty for each group of alleles at a locus j multiplied by the square root of the number of alleles at that site ($\sqrt{p_j}$). Estimation of the HT and IBD allelic effects was done using the `gglasso` R package[41]. Estimated parameters were measured for all LASSO models (See Supplementary section 3).

The last predictive method we tested was the Gaussian Kernel framework proposed by[42]. The primary motivation was to evaluate if the construction of a kernel with multiallelic marker information better captures cryptic epistasis which could benefit genome-wide prediction. The transformed Gaussian Kernel estimates the joint distribution for the bandwidth and form parameters through a Markov Chain Monte Carlo (MCMC) Gibbs sampler. The bandwidth parameter is central in the estimation of the Gaussian Kernel and the form parameter is essentially equivalent to the inverse of the shrinkage parameter used in earlier models. We adapted the scripts provided by[42] with a total MCMC iterations of 77,000 and a burn-in of 2,000. Sampled parameter posterior distributions were also collected for each trait (See Supplementary section 4).

Model evaluation

All model paradigms and predictor types were tested with a simple 5-fold cross-validation procedure with 50 replications. The same folds were used across tested models for valid comparisons across each run. Because we were assessing the accuracy of models which were extensions of the linear model, we used traditional model performance statistics[43]. These include the prediction accuracy (Pearson correlation coefficient), statistical bias (the regression coefficient of the across-trial hybrid BLUP on its predicted value $\beta_{\tilde{y},\hat{y}}$), and the per cent root mean square error of prediction ($100 \times \sqrt{\mathrm{E}\left((\tilde{y}-\hat{y})^2\right)}/\mu$) where each error is scaled by the trait mean. All measures for model performance are provided for each tuber trait studied here (Figure 5.2).

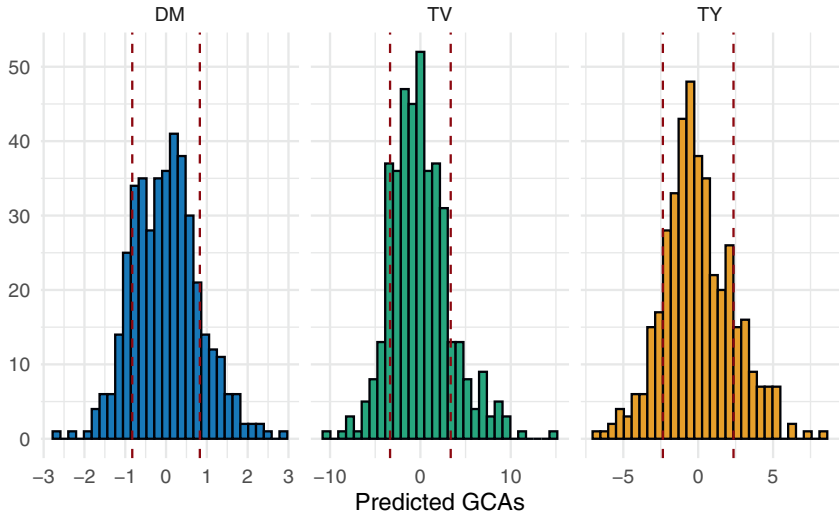
5.3 Results

Using the same BLUPs for the 1,038 hybrids of the multi-environment trial datasets from[33], we were able to estimate the genetic variance and pooled residual variance for these three traits (Figure 5.2). All heritabilities were high with average tuber volume being the highest among the three variates. From this model we were also able to estimate general combining abilities for the 416 parental lines used in the hybrid crossing block. Because some parental lines were not genotyped, a significant fraction of hybrids were excluded from subsequent analysis. Therefore, 783 hybrids were retained for IBD triage and genetic modelling.

The RABBIT pipeline converged providing haplotype profiles for all 416 inbred lines. Founder composition per inbred was quite variable, however, a core of founder backgrounds was observed. Observing the IBD profile for parent line's PL960 and PL314 (bottom and top 1% percentiles for average tuber volume GCA), they appeared to share a common core composed of founder line's FL1, FL2, and FL3. Differential founder contribution between these two inbreds was also observed with large contributions of founder FL5 in PL960 on chromosomes 1, 2 and 5, and large contributions of founder FL4 in

Figure 5.2: The full genetic variance ($2 \cdot \sigma^2_{gca} + \sigma^2_{sca}$), pooled residual variance, and heritability for dry matter content (DM), average tuber volume (TV), and total tuber yield (TY) along with the distribution of predicted general combining abilities (GCAs) for each. The +1/-1 standard deviations of the GCAs are given for each trait.

Trait	σ^2_G	σ^2_ε	H^2
DM	2.49	0.78	0.76
TV	38.31	5.18	0.88
TY	23.80	4.88	0.83



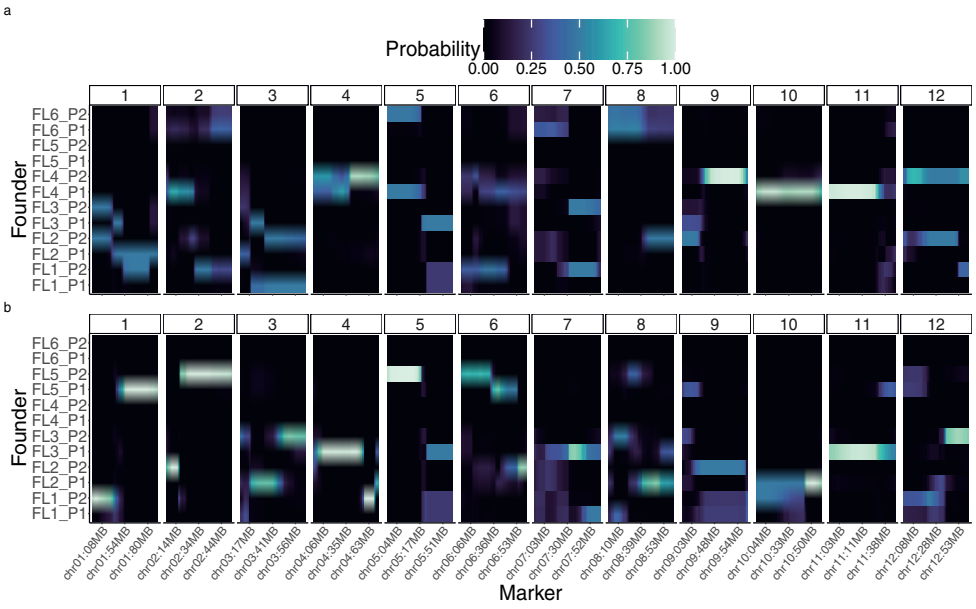


Figure 5.3: The genome-wide identity by descent profiles of (a) the best parental line ("PL314") and (b) the worst parental line ("PL960") with respect to average tuber volume.

PL314 on chromosomes 4, 9, 10, 11, and 12. It is also worth noting the large chromosome blocks with little to no recombination appeared to have occurred since initial founder crossing. This was most apparent on chromosome 11 in both inbreds which inherited two different founder backgrounds.

Looking to predictive performance of our various markerset types, we found that the SNP-based datasets performed the best across traits and were particularly superior using the transformed Gaussian Kernel method (Figure 5.4). Generally, the HT markerset was less performant than SNP and sSNP, but these differences were negligible especially in dry matter content. Additionally, we found very little difference between the ridge regression and LASSO models. The IBD probeset consistently under-performed with respect to prediction accuracy regardless of model or trait tested. We saw minor improvement between the LASSO and grouped LASSO for the IBD predictors, but this was only observed in dry matter content. Considering also the prediction error, the Gaussian Kernel models had consistently the lowest error for all trait predictor combinations with exception of the grouped LASSO in drymatter content (Table 5.2). Differences in predictive bias appeared to be negligible aside from higher rates detected in the LASSO models, but this was generally to be expected.

5.4 Discussion

Examining the first from our set of questions, the biallelic-based predictors (both SNP and sSNP probesets) exhibited the greatest modelling accuracy within our cross-validation

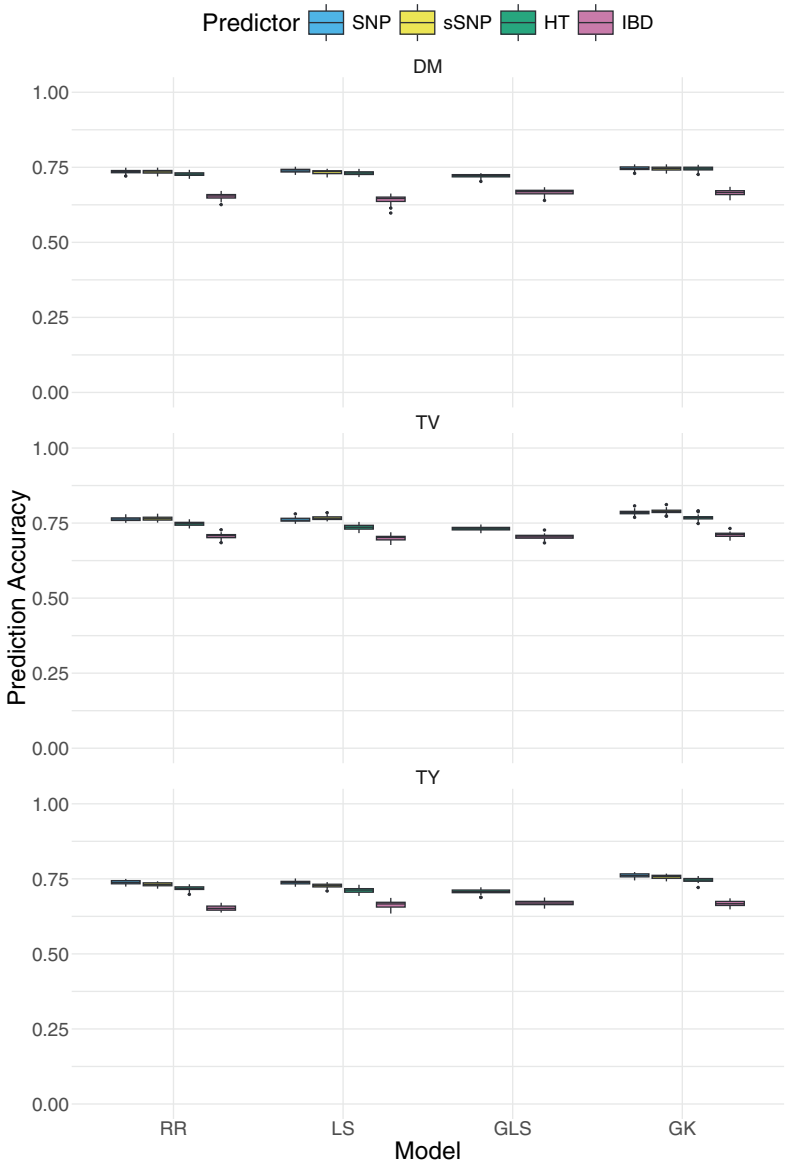


Figure 5.4: Median prediction accuracy for dry matter content (DM), average tuber volume (TV), and total tuber yield (TY) over all models including ridge regression (RR), LASSO (LS), group LASSO (GLS), and the Gaussian Kernel (GK) as well as over each predictor type including short range haplotypes (HT), founder identity by descent probabilities (IBD), SNPs (SNP), and a sparse SNP set (sSNP). The best model within predictor type is given in bold while the best model across predictors is underlined.

Table 5.2: Prediction Accuracy (PA), prediction bias and the per cent root mean square error (RMSE) for dry matter content (DM), average tuber volume (TV), and total tuber yield (TY). This was done over all models including ridge regression (RR), LASSO (LS), group LASSO (GLS), and the Gaussian Kernel (GK) as well as over each predictor type including short range haplotypes (HT), founder identity by descent probabilities (IBD), SNPs (SNP), and a sparse SNP set (sSNP). The best model within predictor type is given in bold while the best model across predictors is underlined.

	DM			TV			TY		
Model	PA	Bias	RMSE	PA	Bias	RMSE	PA	Bias	RMSE
SNP									
RR	0.74	**1.01**	5.78	0.76	**1.00**	16.60	0.74	1.01	22.85
LS	0.74	0.90	5.80	0.76	0.88	16.89	0.74	0.95	22.86
GK	**0.75**	1.02	**5.68**	**0.78**	0.98	**15.90**	**0.76**	**1.00**	**21.97**
sSNP									
RR	0.73	**1.01**	5.80	0.76	**1.00**	16.55	0.73	1.01	23.07
LS	0.74	0.91	5.83	0.77	0.89	16.68	0.73	0.96	23.26
GK	**0.75**	1.01	**5.69**	**0.79**	0.99	**15.77**	**0.76**	**1.00**	**22.15**
HT									
RR	0.73	**1.02**	5.87	0.75	**1.00**	17.09	0.72	**1.00**	23.54
LS	0.73	0.93	5.85	0.74	0.88	17.57	0.71	0.89	23.98
GLS	0.72	0.98	5.91	0.73	0.96	17.50	0.71	0.96	23.97
GK	**0.74**	1.03	**5.71**	**0.77**	0.98	**16.43**	**0.75**	1.01	**22.56**
IBD									
RR	0.65	**0.99**	6.48	0.71	**1.01**	18.12	0.65	**1.00**	25.73
LS	0.65	0.81	6.66	0.70	0.86	18.45	0.67	0.88	25.46
GLS	**0.67**	0.94	**6.36**	0.70	0.91	18.31	**0.67**	0.88	25.36
GK	0.67	1.04	6.38	**0.71**	1.03	**18.02**	0.67	1.05	**25.25**

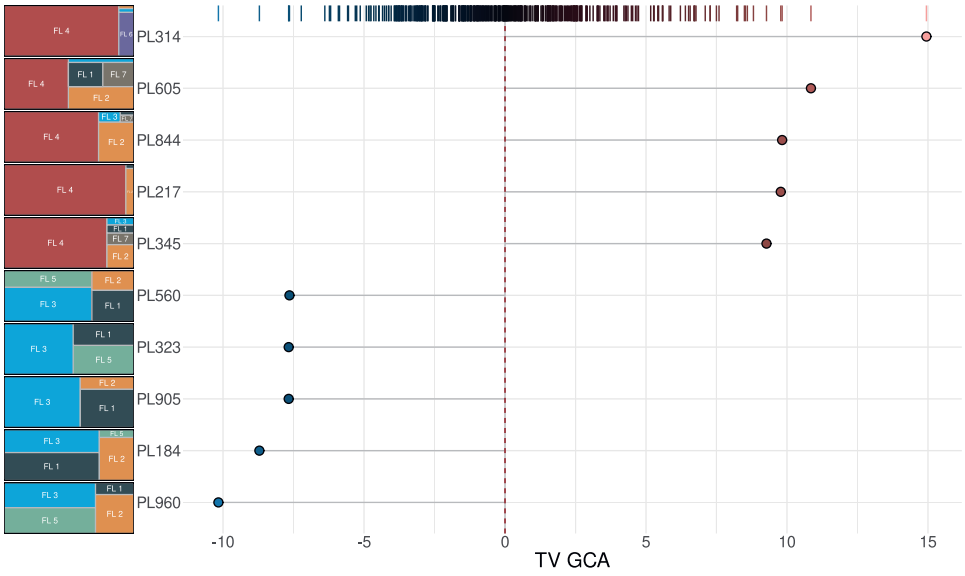


Figure 5.5: The founder origin composition and general combining ability of the top five and worst five performing parental lines with respect to average tuber volume (TV). Distribution of all predicted general combining abilities are provided as a reference.

schema and were stable across statistical models (Figure 5.4). The lack of significant differences between the SNP and sSNP probesets suggest that differences in model performance due to genome coverage are negligible between 2600 and 1700 SNPs. Taking this further, the increased prediction error observed with the IBD predictor set then was not the product of ascertainment bias of filtered regions, but rather the information content of the haplotypes themselves (Table 5.2). In contrast to our IBD set, our haplotag-based set of predictors performed generally better and were nearly identical to the other predictor types; this was most apparent for tuber dry matter content (Figure 5.4). This poses an interesting question of why the HT set is generally identical across models for dry matter content but not so for other traits. One potential explanation could be that the statistical architecture, or *genetic complexity*, of dry matter content is less than that of the other yield components can be generalized well by both the HT and SNP/sSNP probesets in contrast to yield components like tuber volume and gross yield. Additionally, the prediction of tuber volume or total tuber yield could be improved by a haplotag parameterization which considers the relationship of alleles within a locus. One potential solution could to integrate the structural distance between haplotag alleles into a variance-covariance matrix for each haplotype effect[38].

Turning to our second question, there were a few findings with respect to model choice. For our three tuber characteristics, there was very little difference between the ridge regression and LASSO models with exception of a slight increase in prediction bias in the latter (Table 5.2). For almost all trait and predictor types, the Gaussian Kernel performed best minimizing the prediction error with exception of the IBD predictors for dry matter

content. The stable performance of the Gaussian Kernel across traits is worth highlighting as this has been observed in other crops as well. Whether frequentist[44] or Bayesian[45], reproducing Kernel Hilbert spaces are flexible and allow for trait-specific tuning outside a general variance component. Moreover, kernel-based methods are even competitive with deep learning methods which continue to gain popularity in the GP/GS space[46]. We were also particularly interested in whether grouped-penalization methods could better exploit the probe and allele structure in the haplotag and IBD probesets. However, no benefit was observed between the group LASSO and LASSO with exception of the aforementioned IBD predictors for dry matter content. There, the group LASSO performed best with respect to the accuracy and total error. Having said this, the difference observed between the group LASSO and Gaussian Kernel were negligible.

Touching on the final question, IBD-based predictors performed worse than their IBS counterparts in the SNP, sSNP, and HT probesets. To interrogate potential causes, there are multiple explanations. First is the problem of founder choice and which stage of a breeding program you want to consider the *founding* generation. While perhaps obvious to the reader, there should be sufficient marker variation (i.e. variation via IBS) between founders in order to both estimate IBD-based origin and in order for it to be useful in genetic applications. Simulation studies confirm that the benefits of IBD-based marker effects are most pronounced when there is sufficient marker diversity in the founding population[21]. Not only does marker variation need to be present, but there must also be relevant *genetic* variation for the model to pick up on. Studies in barley have shown that when the QTL effects are nearly identical among founders, modelling on founder effects will not yield benefits over that of simpler IBS formulations[47,48]. So in order to maximize the utility and accuracy of IBD information in a breeding program, we would need to identify the generation which offers the greatest genetic contrast among some set of intermediate genotypes. There could be alternatives to this through certain pre-processing methods such as structuring founder effects by the genetic relationship of the founders (via IBS or IBD) or even transforming these effects through spectral decomposition as recommended by[29]. However, this leads to issues of interpretability of the estimated parameters. Several population-specific factors such as a early selection intensities might also be related to the underperformance of the IBD probeset in our genome-wide prediction scenario.

Regardless of whether multiallelic data are used in a statistical model, its utility in a breeding program is not bound there. Extending our example from Figure 5.3 to include the 5^{th} and 95^{th} percentiles for tuber volume GCA, we can examine their founder line composition (Figure 5.5). Evidently, those parental lines with larger tuber volume GCAs tend to be over-represented by the founder FL4 while those parental lines with the smallest tuber volume GCA appear to share large compositions of founders FL3 and FL5. While the IBD information was not able to offer any benefits in the context of GP/GS, this does not mean it is unable to express insight to a breeder regarding their material. Genome-wide prediction is crucial for informed selection and advancement of germplasm, but there is still information which influences decision making even if it does not enter the linear mixed model. Activities like breeding strategy development (e.g. maintaining breeding

pool genetic variance) or breeding cycle evaluation (e.g. evaluating the impact of previous selection cycles) could both benefit with the introduction of IBD information.

5.5 Conclusion

As potato breeding progresses through the *big-data* era, it is crucial for geneticists and breeders to assess which information best informs them of a candidate's genetic potential. With regard to molecular technologies, the primary questions ought to focus on adequacy of a marker platform for a specific segment or program along with whether it is an appropriate choice given the crop's genomic complexity. Next then, is how that technology could best be leveraged to predict genetic value, whether it be a simple monogenic target like biotic resistance or a traditionally complex trait such as yield fractions. The key findings of this paper suggest that SNP variants are sufficient for genome-wide prediction in hybrid potato. The use of multi-allelic marker information such as our haplotag probeset are also possible and offer a unified prediction approach with a powerful genotyping strategy. IBD-based predictors also succeeded in the prediction of hybrid yield, though with a noticeable penalty. Future work should consider further modelling strategies and alternative parameterisations for multi-allelic probesets (whether IBS or IBD) to ensure marker information is used efficiently.

Bibliography

1. Mendoza HA and Haynes FL. Genetic Basis of Heterosis for Yield in the Autotetraploid Potato. Springer. 1974; 45:21–5
2. Tai GCC. Analysis of Genotype-Environment Interactions of Potato Yield[1]. Crop Science. 1979 Jul; 19:434–8. DOI: 10.2135/cropsci1979.0011183X001900040003x. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci1979.0011183X001900040003x [Accessed on: 2025 Jan 3]
3. Almekinders CJM, Chujoy E, and Thiele G. The Use of True Potato Seed as Propoor Technology: The Efforts of an International Agricultural Research Institute to Innovating Potato Production. Potato Research. 2009 Nov; 52:275–93. DOI: 10.1007/s11540-009-9142-5. Available from: https://doi.org/10.1007/s11540-009-9142-5 [Accessed on: 2025 Jan 3]
4. Krantz FA. Potato Breeding Methods. Vol. 25. Technical Bulletin. University Farm, St. Paul: University of Minnesota, 1924 Dec
5. Swaminathan MS. Nature of Polyploidy in Some 48-Chromosome Species of the Genus Solanum, Section Tuberarium. Genetics. 1954; 39:59. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209637/ [Accessed on: 2024 Dec 29]
6. Lian Q, Tang D, Bai Z, Qi J, Lu F, Huang S, and Zhang C. Acquisition of Deleterious Mutations during Potato Polyploidization. Journal of Integrative Plant Biology. 2019 Jan 1; 61:7–11. DOI: 10.1111/jipb.12748
7. Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B, Yang X, Zeng Z, Douches DS, Jiang J, Veilleux RE, and Buell CR. Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated Solanum Tuberosum. The Plant cell. 2016 Feb; 28:388–405. DOI: 10.1105/tpc.15.00538. PMID: 26772996. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26772996 [Accessed on: 2019 Feb 14]
8. Marand AP, Jansky SH, Gage JL, Hamernik AJ, de Leon N, and Jiang J. Residual Heterozygosity and Epistatic Interactions Underlie the Complex Genetic Architecture of Yield in Diploid Potato. Genetics. 2019 Mar 18 :genetics.302036.2019. DOI: 10.1534/genetics.119.302036. PMID: 30885982. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30885982 [Accessed on: 2019 Apr 25]
9. Kyriakidou M, Achakkagari SR, Gálvez López JH, Zhu X, Tang CY, Tai HH, Anglin NL, Ellis D, and Strömvik MV. Structural Genome Analysis in Cultivated Potato Taxa. Theoretical and Applied Genetics. 2020 Mar; 133:951–66. DOI: 10.1007/

s00122-019-03519-6. Available from: https://doi.org/10.1007/s00122-019-03519-6 [Accessed on: 2024 Dec 8]

10. The Potato Genome Sequencing Consortium. Genome Sequence and Analysis of the Tuber Crop Potato. Nature. 2011 Jul; 475:189–95. DOI: 10.1038/nature10158. Available from: https://www.nature.com/articles/nature10158 [Accessed on: 2024 Dec 29]
11. Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, Li D, Zhu G, Wang H, Zhou Y, Zhou Y, Bryan GJ, Buell CR, Zhang C, and Huang S. Genome Evolution and Diversity of Wild and Cultivated Potatoes. Nature. 2022 Jun; 606:535–41. DOI: 10.1038/s41586-022-04822-x. Available from: https://www.nature.com/articles/s41586-022-04822-x [Accessed on: 2025 Jan 3]
12. Kloosterman B, Abelenda JA, Gomez MdMC, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, Visser RGF, and Bachem CWB. Naturally Occurring Allele Diversity Allows Potato Cultivation in Northern Latitudes. Nature. 2013 Mar; 495:246–50. DOI: 10.1038/nature11912. Available from: https://search.ebscohost.com/login.aspx?direct=true&db=pbh&AN=86049456&site=ehost-live [Accessed on: 2024 Nov 20]
13. Eck HJV, Oortwijn ME, Terpstra IR, Lieshout NHV, Knaap EVD, Willemsen JH, and Bachem CW. Engineering of Tuber Shape in Potato (Solanum Tuberosum) with Marker Assisted Breeding or Genetic Modification Using StOFP20. preprint. In Review, 2022 Jul. DOI: 10.21203/rs.3.rs-1807189/v1. Available from: https://www.researchsquare.com/article/rs-1807189/v1 [Accessed on: 2024 Jan 4]
14. Ames M, Hamernik A, Behling W, Douches DS, Halterman DA, and Bethke PC. A Survey of the Sli Gene in Wild and Cultivated Potato. Plant Direct. 2024; 8:e589. DOI: 10.1002/pld3.589. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/pld3.589 [Accessed on: 2024 Nov 20]
15. Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, and Van Eck HJ. Development and Analysis of a 20K SNP Array for Potato (Solanum Tuberosum): An Insight into the Breeding History. Theoretical and Applied Genetics. 2015 Dec; 128:2387–401. DOI: 10.1007/s00122-015-2593-y. Available from: http://link.springer.com/10.1007/s00122-015-2593-y [Accessed on: 2025 Jan 3]
16. Meijer D, Viquez-Zamora M, Van Eck HJ, Hutten RCB, Su Y, Rothengatter R, Visser RGF, Lindhout WH, and Van Heusden AW. QTL Mapping in Diploid Potato by Using Selfed Progenies of the Cross S. Tuberosum × S. Chacoense. Euphytica. 2018 Jul; 214:121. DOI: 10.1007/s10681-018-2191-6. Available from: https://link.springer.com/10.1007/s10681-018-2191-6 [Accessed on: 2025 Jan 3]
17. Wilson S, Malosetti M, Maliepaard C, Mulder HA, Visser RGF, and van Eeuwijk F. Training Set Construction for Genomic Prediction in Auto-Tetraploids: An Example in Potato. Frontiers in Plant Science. 2021; 12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.771075 [Accessed on: 2022 Nov 15]
18. Endelman JB, Kante M, Lindqvist-Kreuze H, Kilian A, Shannon LM, Caraza-Harter MV, Vaillancourt B, Mailloux K, Hamilton JP, and Buell CR. Targeted

Genotyping-by-Sequencing of Potato and Data Analysis with R/polyBreedR. The Plant Genome. 2024; 17:e20484. DOI: 10.1002/tpg2.20484. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/tpg2.20484 [Accessed on: 2024 Nov 10]

19. Leyva-Pérez MdlO, Vexler L, Byrne S, Clot CR, Meade F, Griffin D, Ruttink T, Kang J, and Milbourne D. PotatoMASH—A Low Cost, Genome-Scanning Marker System for Use in Potato Genomics and Genetics Applications. Agronomy. 2022 Oct; 12:2461. DOI: 10.3390/agronomy12102461. Available from: https://www.mdpi.com/2073-4395/12/10/2461 [Accessed on: 2024 Sep 3]
20. Tinker NA, Bekele WA, and Hattori J. Haplotag: Software for Haplotype-Based Genotyping-by-Sequencing Analysis. G3 Genes\textbarGenomes\textbarGenetics. 2016 Apr; 6:857–63. DOI: 10.1534/g3.115.024596. Available from: https://doi.org/10.1534/g3.115.024596 [Accessed on: 2024 Oct 13]
21. Thérèse Navarro A, Tumino G, Voorrips RE, Arens P, Smulders MJM, van de Weg E, and Maliepaard C. Multiallelic Models for QTL Mapping in Diverse Polyploid Populations. BMC Bioinformatics. 2022 Dec; 23:1–16. DOI: 10.1186/s12859-022-04607-z. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04607-z [Accessed on: 2025 Jan 2]
22. Clot CR, Wang X, Koopman J, Navarro AT, Bucher J, Visser RGF, Finkers R, and Van Eck HJ. High-Density Linkage Map Constructed from a Skim Sequenced Diploid Potato Population Reveals Transmission Distortion and QTLs for Tuber Yield and Pollen Shed. Potato Research. 2024 Mar; 67:139–63. DOI: 10.1007/s11540-023-09627-7. Available from: https://link.springer.com/10.1007/s11540-023-09627-7 [Accessed on: 2024 May 7]
23. Vexler L, Leyva-Perez MdlO, Konkolewska A, Clot CR, Byrne S, Griffin D, Ruttink T, Hutten RCB, Engelen C, Visser RGF, Prigge V, Wagener S, Lairy-Joly G, Driesprong JD, Riis Sundmark EH, Rookmaker ANO, van Eck HJ, and Milbourne D. QTL Discovery for Agronomic and Quality Traits in Diploid Potato Clones Using PotatoMASH Amplicon Sequencing. G3 Genes\textbarGenomes\textbarGenetics. 2024 Oct; 14:jkae164. DOI: 10.1093/g3journal/jkae164. Available from: https://doi.org/10.1093/g3journal/jkae164 [Accessed on: 2024 Nov 10]
24. Falconer DS and Mackay TFC. Introduction to Quantitative Genetics. Longman, 1996. 464 pp.
25. Li W, Boer MP, Zheng C, Joosen RVL, and van Eeuwijk FA. An IBD-based Mixed Model Approach for QTL Mapping in Multiparental Populations. Theoretical and Applied Genetics 2021 134:11. 2021 Aug; 134:3643–60. DOI: 10.1007/S00122-021-03919-7. Available from: https://link.springer.com/article/10.1007/s00122-021-03919-7 [Accessed on: 2021 Nov 1]
26. Garin V, Wimmer V, Borchardt D, Malosetti M, and Van Eeuwijk F. The Influence of QTL Allelic Diversity on QTL Detection in Multi-Parent Populations: A Simulation Study in Sugar Beet. BMC Genomic Data. 2021 Dec; 22:4. DOI: 10.1186/s12863-021-00960-9. Available from: https://bmcgenet.biomedcentral.com/articles/10.1186/s12863-021-00960-9 [Accessed on: 2025 Jan 4]

27. Zheng C, Boer MP, and van Eeuwijk FA. Reconstruction of Genome Ancestry Blocks in Multiparental Populations. 2015. DOI: 10.1534/genetics.115.177873. Available from: www.genetics.org/lookup/suppl/ [Accessed on: 2018 Nov 26]
28. Tisné S, Denis M, Cros D, Pomiès V, Riou V, Syahputra I, Omoré A, Durand-Gasselin T, Bouvet JM, and Cochard B. Mixed Model Approach for IBD-based QTL Mapping in a Complex Oil Palm Pedigree. BMC Genomics. 2015 Oct; 16:798. DOI: 10.1186/s12864-015-1985-3. Available from: https://doi.org/10.1186/s12864-015-1985-3 [Accessed on: 2025 Jan 3]
29. Van Eeuwijk FA, Boer M, Totir LR, Bink M, Wright D, Winkler CR, Podlich D, Boldman K, Baumgarten A, Smalley M, Arbelbide M, ter Braak CJF, and Cooper M. Mixed Model Approaches for the Identification of QTLs within a Maize Hybrid Breeding Program. Theoretical and Applied Genetics. 2010 Jan; 120:429–40. DOI: 10.1007/s00122-009-1205-0. Available from: http://link.springer.com/10.1007/s00122-009-1205-0 [Accessed on: 2019 Jul 16]
30. Hill WG. Applications of Population Genetics to Animal Breeding, from Wright, Fisher and Lush to Genomic Prediction. Genetics. 2014 Jan; 196:1–16. DOI: 10.1534/genetics.112.147850. Available from: https://doi.org/10.1534/genetics.112.147850 [Accessed on: 2025 Jan 4]
31. Korontzis G, Malosetti M, Zheng C, Maliepaard C, Mulder HA, Lindhout P, Veerkamp RF, and van Eeuwijk FA. QTL Detection in a Pedigreed Breeding Population of Diploid Potato. Euphytica. 2020 Aug; 216:145. DOI: 10.1007/s10681-020-02674-y. Available from: https://doi.org/10.1007/s10681-020-02674-y [Accessed on: 2023 Jan 3]
32. Amadeu RR, Muñoz PR, Zheng C, and Endelman JB. QTL Mapping in Outbred Tetraploid (and Diploid) Diallel Populations. Genetics. 2021 Aug; 219:iyab124. DOI: 10.1093/genetics/iyab124. PMID: 34740237. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8570786/ [Accessed on: 2023 Aug 22]
33. Adams J, De Vries M, and Van Eeuwijk F. Efficient Genomic Prediction of Yield and Dry Matter in Hybrid Potato. Plants. 2023 Jul 11; 12:2617. DOI: 10.3390/plants12142617. Available from: https://www.mdpi.com/2223-7747/12/14/2617
34. Lindhout P, Meijer D, Schotte T, Hutten RC, Visser RG, and Eck HJ van. Towards F1 hybrid seed potato breeding. Potato Research 2011; 54:301–12
35. LGC. SeqSNP Targeted GBS as Alternative for Array Genotyping in Routine Breeding Programs [White Paper]. Biosearch Technologies, 2019. Available from: https://biosearch-cdn.azureedge.net/assetsv6/seqsnp-tgbs-alternative-genotyping-routine-breeding-programs.pdf
36. Heslot N, Rutkoski J, Poland J, Jannink JL, and Sorrells ME. Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. PLOS ONE. 2013 Sep; 8:e74612. DOI: 10.1371/journal.pone.0074612. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074612 [Accessed on: 2024 Dec 22]
37. Da Y. Multi-Allelic Haplotype Model Based on Genetic Partition for Genomic Prediction and Variance Component Estimation Using SNP Markers. BMC Genetics. 2015 Dec; 16:1–12. DOI: 10.1186/s12863-015-0301-1. Available from: https:

//bmcgenomdata.biomedcentral.com/articles/10.1186/s12863-015-0301-1 [Accessed on: 2023 Sep 14]

38. Schaid DJ. Evaluating Associations of Haplotypes with Traits. Genetic Epidemiology. 2004 Dec; 27:348–64. DOI: 10.1002/gepi.20037. Available from: https://onlinelibrary.wiley.com/doi/10.1002/gepi.20037 [Accessed on: 2024 Nov 21]
39. Friedman JH, Hastie T, and Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software. 2010 Feb; 33:1–22. DOI: 10.18637/jss.v033.i01. Available from: https://doi.org/10.18637/jss.v033.i01 [Accessed on: 2024 Oct 3]
40. Ogutu JO and Piepho HP. Regularized Group Regression Methods for Genomic Prediction: Bridge, MCP, SCAD, Group Bridge, Group Lasso, Sparse Group Lasso, Group MCP and Group SCAD. BMC Proceedings. 2014 Oct; 8:S7. DOI: 10.1186/1753-6561-8-S5-S7. Available from: https://doi.org/10.1186/1753-6561-8-S5-S7 [Accessed on: 2024 Sep 2]
41. Yang Y and Zou H. A Fast Unified Algorithm for Solving Group-Lasso Penalize Learning Problems. Statistics and Computing. 2015 Nov; 25:1129–41. DOI: 10.1007/s11222-014-9498-5. Available from: https://doi.org/10.1007/s11222-014-9498-5 [Accessed on: 2024 Oct 3]
42. Pérez-Elizalde S, Cuevas J, Pérez-Rodríguez P, and Crossa J. Selection of the Bandwidth Parameter in a Bayesian Kernel Regression Model for Genomic-Enabled Prediction. Journal of Agricultural, Biological, and Environmental Statistics. 2015 Dec; 20:512–32. DOI: 10.1007/s13253-015-0229-y. Available from: https://doi.org/10.1007/s13253-015-0229-y [Accessed on: 2024 Aug 23]
43. Hsu WL, Garrick DJ, and Fernando RL. The Accuracy and Bias of Single-Step Genomic Prediction for Populations Under Selection. G3 Genes\textbarGenomes\textbarGenetics. 2017 Aug; 7:2685–94. DOI: 10.1534/g3.117.043596. Available from: https://doi.org/10.1534/g3.117.043596 [Accessed on: 2024 Dec 22]
44. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome. 2011; 4. DOI: 10.3835/plantgenome2011.08.0024. Available from: https://onlinelibrary.wiley.com/doi/abs/10.3835/plantgenome2011.08.0024 [Accessed on: 2022 Dec 9]
45. Gianola D and van Kaam JBCHM. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics. 2008 Apr; 178:2289–303. DOI: 10.1534/genetics.107.084285. Available from: https://doi.org/10.1534/genetics.107.084285 [Accessed on: 2022 Dec 9]
46. Crossa J, Martini JWR, Gianola D, Pérez-Rodríguez P, and Jarquin D. Deep Kernel and Deep Learning for Genome-Based Prediction of Deep Kernel and Deep Learning for Genome-Based Prediction of Single Traits in Multienvironment Breeding Trials Single Traits in Multienvironment Breeding Trials. Frontiers in Genetics. 2019; 10:1–13. DOI: 10.3389/fgene.2019.01168. Available from: https://digitalcommons.unl.edu/agronomyfacpub [Accessed on: 2020 Jan 7]

47. Maurer A, Draba V, and Pillen K. Genomic Dissection of Plant Development and Its Impact on Thousand Grain Weight in Barley through Nested Association Mapping. Journal of Experimental Botany. 2016 Apr; 67:2507–18. DOI: 10.1093/jxb/erw070. PMID: 26936829
48. Maurer A, Sannemann W, Léon J, and Pillen K. Estimating Parent-Specific QTL Effects through Cumulating Linked Identity-by-State SNP Effects in Multiparental Populations. Heredity. 2017 May; 118:477–85. DOI: 10.1038/hdy.2016.121. Available from: https://www.nature.com/articles/hdy2016121 [Accessed on: 2025 Jan 4]

5.6 Supplementary Material

Founder Origin Sparsity

After the conclusion of the RABBIT pipeline, founder origin sparsity was assessed genome-wide using a rolling-window over average number of sites which had no confident haplotype call (probability under 0.5). This showed a high degree of missing haplotype calls on chromosome's 1 and 5 (Sup Fig S1).

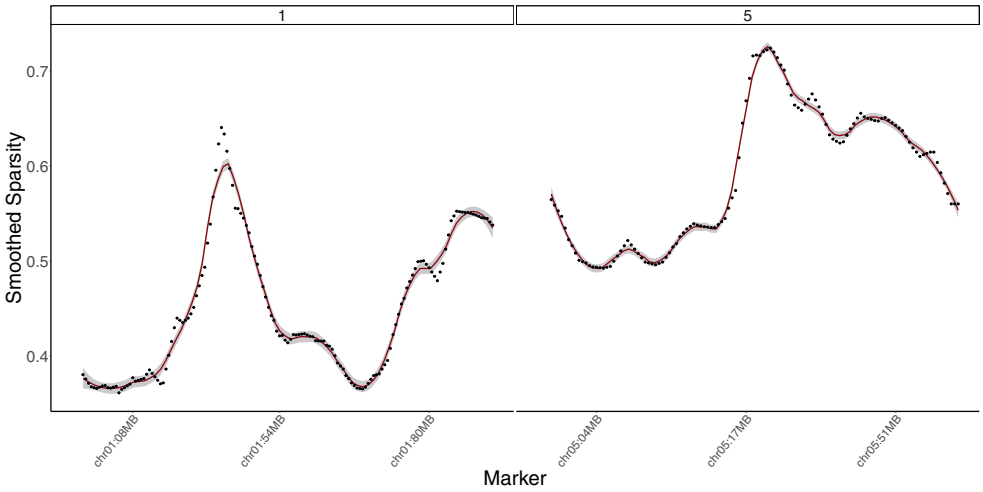


Figure S1: Smoothed sparsity of founder-origin alleles for chromosome's 1 and 5.

Multiallelic Marker Matrix Parameterization

For both haplotag and IBD based markersets, let us assign each allele some identifier {a, b, c} to denote genotypic states with the former representing a haplotype sequence of 75 basepairs in length and the latter being an assigned founder origin. We then consider two loci (l1 and l2) with the reference allele (a) being either the reference haplotag allele or the most prevalent founder allele in the population. The multiallelic incidence matrix is centred using the cross product of a column vector of 1's $\left(\mathbf{1}_4^T = (1,1,1,1)^T\right)$ and a vector of allelic frequencies excluding the reference allele $\left(\mathbf{p} = \begin{bmatrix} l1b & l1c & l2b \\ \frac{3}{8} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}\right)$.

$$
\begin{array}{c} \\ g1 \\ g2 \\ g3 \\ g4 \end{array}
\begin{array}{c} \begin{array}{cc} l1 & l2 \end{array} \\ \begin{bmatrix} aa & ab \\ cc & bb \\ bb & aa \\ ab & ab \end{bmatrix} \end{array}
\Rightarrow
\begin{array}{c} \\ g1 \\ g2 \\ g3 \\ g4 \end{array}
\begin{array}{c} \begin{array}{ccccc} l1a & l1b & l1c & l2a & l2b \end{array} \\ \begin{bmatrix} 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 2 \\ 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix} \end{array}
\Rightarrow
\begin{array}{c} \\ g1 \\ g2 \\ g3 \\ g4 \end{array}
\begin{array}{c} \begin{array}{ccc} l1b & l1c & l2b \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 2 \\ 2 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \end{array}
\Rightarrow
$$

$$\mathbf{1}_4^T \bigotimes 2\mathbf{p} - \begin{array}{c} \\ g1 \\ g2 \\ g3 \\ g4 \end{array} \begin{array}{c} \begin{array}{ccc} l1b & l1c & l2b \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 2 \\ 2 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \end{array} \Rightarrow \begin{array}{c} \\ g1 \\ g2 \\ g3 \\ g4 \end{array} \begin{array}{c} \begin{array}{ccc} l1b & l1c & l2b \end{array} \\ \begin{bmatrix} 2p_{l1b} & 2p_{l1c} & 2p_{l2b}-1 \\ 2p_{l1b} & 2p_{l1c}-2 & 2p_{l2b}-2 \\ 2p_{l1b}-2 & 2p_{l1c} & 2p_{l2b} \\ 2p_{l1b}-1 & 2p_{l1c} & 2p_{l2b}-1 \end{bmatrix} \end{array} \Rightarrow$$

$$\begin{array}{c} \\ g1 \\ g2 \\ g3 \\ g4 \end{array} \begin{array}{c} \begin{array}{ccc} l1b & l1c & l2b \end{array} \\ \begin{bmatrix} \frac{3}{4} & \frac{1}{2} & 0 \\ \frac{3}{4} & -\frac{3}{2} & -1 \\ -\frac{5}{4} & \frac{1}{2} & 1 \\ -\frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

This centred matrix of either IBD founder or haplotag alleles was then used in subsequent genome-wide regression models.

LASSO and GROUP LASSO

For the LASSO (LS) training runs, the degrees of freedom and estimated λ parameter were recorded. This is shown for all traits in (Sup Fig S2). For the grouped LASSO (GLS), the predictive accuracy of that training run is also captured and shown only for the IBD probeset in (Sup Fig S3).

Sampled Posterior Distributions for Gaussian Kernel parameters

The Markov Chain Monte Carlo estimation of our parameters for the Transformed Gaussian Kernel are shown over all runs in (Sup Fig S4) and (Sup Fig S5). These represent the sampled joint posterior distributions of the bandwidth and form parameters $((\hat{h},\ \hat{\phi}))$ and the estimated genetic and residual variances $((\hat{\sigma}_g, \tilde{\hat{\sigma}}_e))$, respectively.

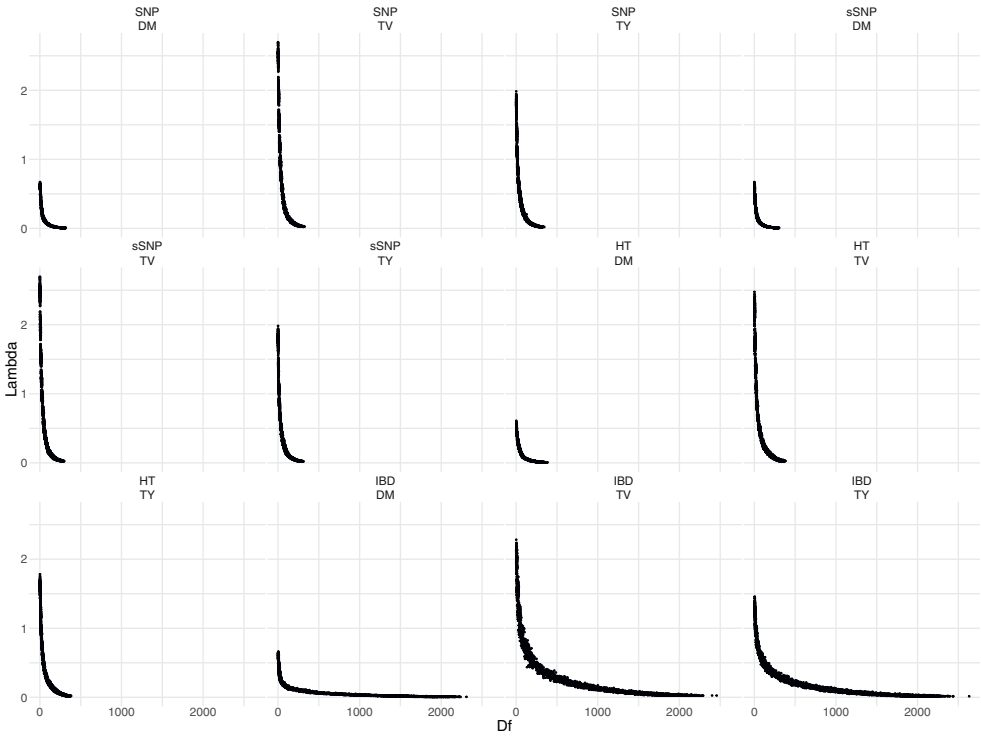


Figure S2: The estimated λ with their associated degrees of freedom for each predictor set and trait for all lasso models.

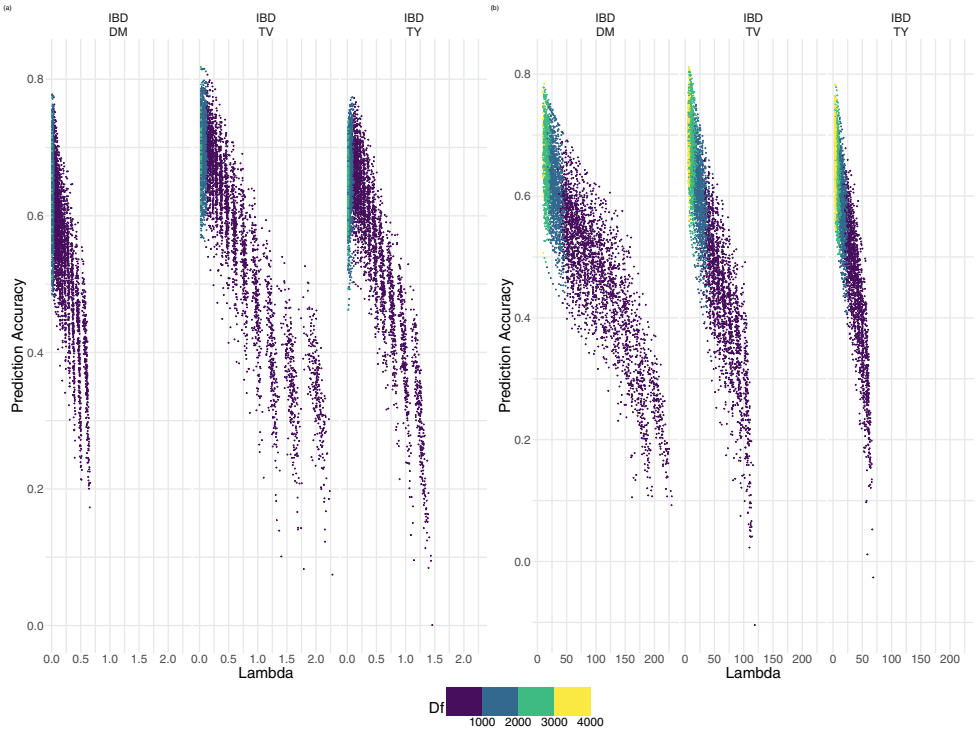


Figure S3: The estimated λ plotted against the model prediction accuracy with the associated degrees of freedom for the multiallelic predictor sets for traditional (a) and (b) grouped lasso models using IBD-based predictors.

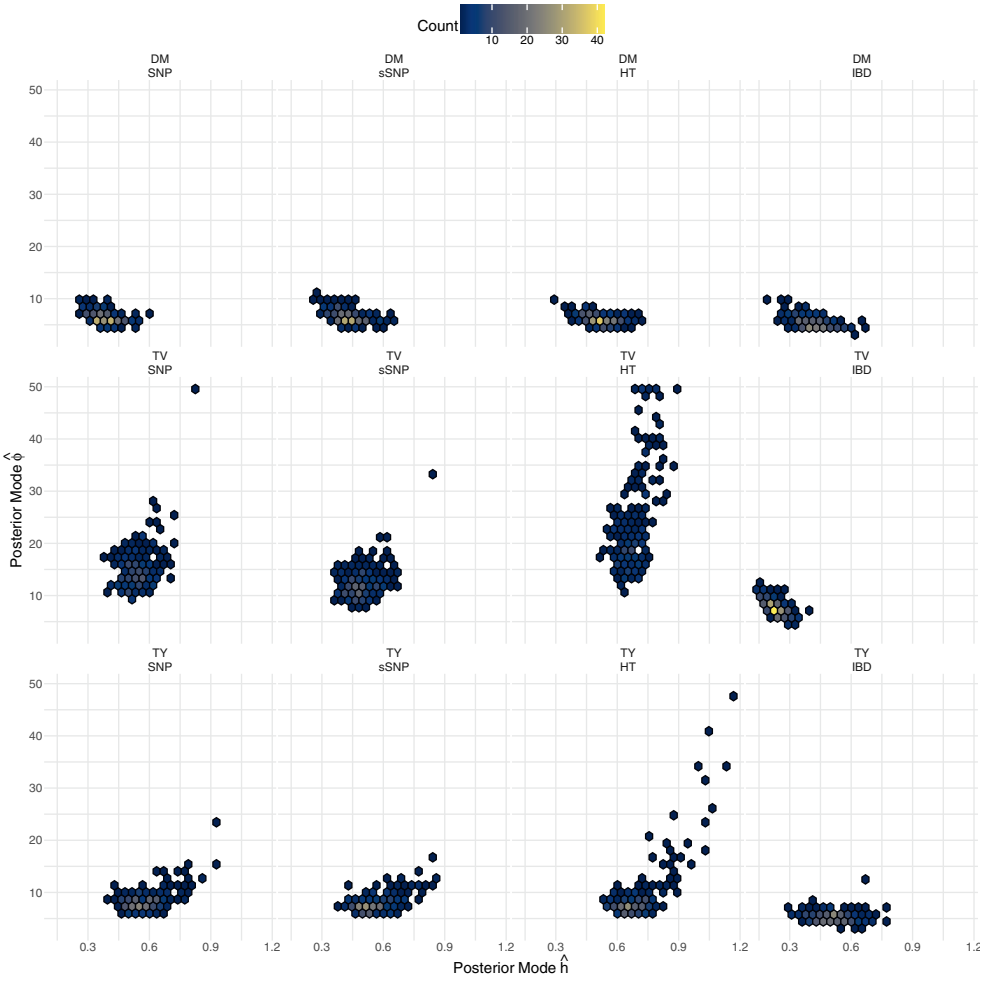


Figure S4: Bivariate heat map of the posterior mode of $\hat{h}$ and $\hat{\phi}$ from each fold of the cross-validation for each trait.

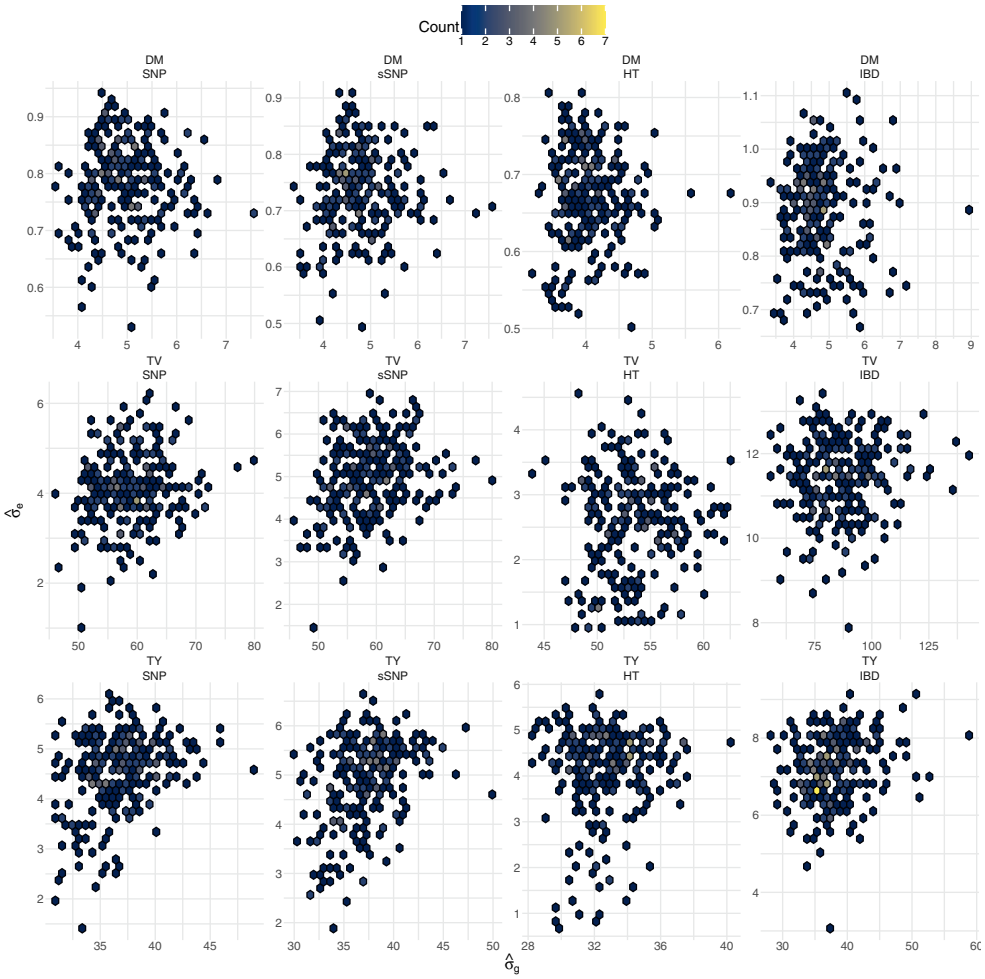


Figure S5: Bivariate heat map of the posterior mode of $\hat{\sigma}_g$ and $\hat{\sigma}_e$ from each fold of the cross-validation for each trait.

6

General Discussion

6.1 Introduction

Hastened genetic progress in potato is dependent on the continued development and application of quantitative genetics methodology in potato breeding. Like many vegetable crops, the application of statistical genetic theory has only received significant attention the past two decades in potato. Before this, complex trait improvement in conventional potato breeding was reliant on phenotypic selection and marked by several inefficiencies including underpowered field evaluations and long breeding cycles[1]. With the continued application of technologies that have had positive success in other field crops, increased genetic improvement in potato is possible. These include both methods such as variance component estimation, genomic prediction, selection index theory as well as the importing of other breeding schemas like recurrent selection or line breeding.

In this thesis, I consider the conversion of clonal tetraploid potato to a diploid hybrid crop as an exercise of applied biometrical methods. To our knowledge, this thesis is the first to interrogate the genetic architecture of a large F1 crossing block in diploid hybrid potato. Moreover, these past experimental chapters also took the opportunity to assess the adequacy of multiple statistical methods in the estimation and prediction of genetic value in this novel crop. I first examined the genetical properties of the F1 hybrid seedling trials including genetic correlations between multiple tuber yield components (chapter 2;[2]). I also utilized multiple forms of genetic information to assess the family history of the inbred parental lines. This included the use of pedigree and molecular marker information (chapter 3;[3]) as well as identity by descent profiles (chapter 5) to evaluate the population structure of these parental lines. After having estimated these genetic parameters, I tested multiple predictive modelling methodologies testing the potential benefits of predicting dominance effects (chapter 3;[3]), the utility of marker assisted selection for complex traits (chapter 4), and various multiallelic parameterisations for genome-wide regression (chapter 5). These topics cover an important range of topics for modern selection methods in a potato breeding program.

Here I provide a synthesis of the previous four experimental chapters with two major aims. I first wish to identify the most important findings of this research, and second, outline the direct implications of our research on potato breeding more generally. I then conclude this thesis with reflections on future research.

Biometrical analysis of field trials

A core activity in breeding programs is the regular evaluation of new parental combinations to screen for genetically improved offspring. Identification of suitable and informative trialling locations ensures an unbiased and representative assessment of a candidate's performance. For this reason, within-trial and multi-environmental trial design is so heavily emphasized in modern crop literature for sufficient statistical power and optimal resource allocation[4]. This permits proper extraction of spatial and other non-genetic components in a field trial and unbiased evaluation of genotype performance. From this successful spatial analysis, it was then possible to measure the genetic parameters of interest whether it be comparison of genotype means, or the estimation of genetic correlations between

traits.

In conventional potato breeding, formal designed trials are reserved for the 3rd or 4th clonal generations when enough clonal propagule can guarantee sufficient replication[5,6]. This is in contrast to other field crops where true seed is the predominant source of germplasm and allows for suitable replication after the initial cross or selfing[7]. An overlooked property of hybrid breeding in potato is the potential of leveraging true seed and the ability to conduct trialing on multi-plant plots over traditional tuber-sown hills[8,9]. The latter tend to be underpowered, with insufficient precision to support early selection. Thus, true potato seed could be a powerful alternative resource in breeding germplasm evaluation. [10,11]. This is not without negative consequences, chief among them, being the poor correlation in multiple tuber traits between the seedling and clonal generations[12–14]. This is addressed in full in Section 6.2, but suffice it to say that the field evaluation of hybrid potato via seedling transplants provides opportunities for explicit field design with sufficiently high precision for multiple tuber phenotypes[15].

In **chapter 2**, such an exercise was conducted in the form of a multi-trial analysis on several hundred F1 hybrids. A penalized splines based procedure was used for spatially de-trending our hybrid observations for three tuber phenotypes. These de-trended phenotypes were then placed in a multivariate linear mixed model for the estimation of the genetic (co-)variances for the GCA, SCA, and several interactive effects. In **chapter 3**, this was followed up with a traditional linear mixed modelling procedure where multiple row column designs with various residual structures were tested on an augmented dataset[16]. The BLUEs and standard errors from these models where then used in a full genomic model where genomically estimated BLUPs were evaluated in a cross-validation schema. While the spatial modelling procedures from **chapter 2** and **chapter 3** were derived from two different frameworks, little to no difference was found between the use of semi-parametric Spline functions and structuring our residual variance. Previous studies have also suggested these approaches as equivalent with exception of the ease of a single model-fitting over the testing of multiple linear mixed models[17]. Both modelling procedures from these chapters follow the general form of a two-stage mixed model analysis where trials are independently analyzed followed by an combined analysis on the genotypic means (most commonly) together with weights on those genotypic observations[18]. Two-stage approaches are understood to be generally less accurate than their single stage counterparts especially with greater model complexity[19]. All this being said, single-stage approaches tend to be unwieldy in their specification and computation and the consequences of a two-stage approach are unlikely to change any inferential conclusions from these two chapters[20].

A core concept in hybrid breeding is the ability to assess the potential of hybrid crosses on the basis of parental value. In the case for phenotypic selection, general combining abilities tend to be the primary mode which communicates a line's value[21]. In **chapter 2**, GCAs were generated for 400 parental lines in multiple tuber traits and found that the GCAs sufficiently captured between 59 and 71 per cent of the observed phenotypic variation in the F1 screening trials. The use of GCA for selection is not unique to hybrid crops with many examples of their use in tetraploid potato in many modern applica-

tions as well[22,23]. This is despite many tetraploid studies reporting a large proportion of non-additive genetic effects controlling tuber qualities like tuber size, tuber yield, and marketable yield[24–26].

From **chapter 2**, some insight around the genetic architecture of yield components was gleaned in the form of genetic correlations. This was not just expressed in terms of a single genetic component over the hybrids, but explicitly in terms of the additive and non-additive genetic covariance ($\Sigma_G = 2 \cdot \Sigma_{gca} + \Sigma_{sca}$). This is worth re-emphasis as there are many complex traits in potato which must undergo selection together in a breeding program [27]. The capacity to decompose the relationships between these correlated traits offers insight to breeders. Additionally, hybrid breeding schemas could benefit from such information to aid in breeding schema design and pool development strategy. Multivariate modelling can also augment genomic prediction in hybrid breeding schemas making it a valuable tool both for inference and predictive applications[28].

Genomic prediction in hybrid potato

Genomic prediction has emerged as a cornerstone of modern crop breeding, offering unprecedented precision in estimating genetic merit by leveraging genome-wide marker data. By modelling molecular markers with phenotypic records, genomic prediction enables breeders to identify superior parental combinations and hybrid crosses with greater accuracy and efficiency than traditional pedigree-based methods. This section explores the application, advantages, and limitations of genomic prediction in hybrid potato breeding, drawing on insights from our experimental chapters to explore its role in accelerating genetic gain in potato.

Multiple chapters in this thesis addressed genomic prediction applications in hybrid potato. In **chapter 3**, the genetic models partitioned into GCA and SCA from **chapter 2** were built upon with molecular marker information to structure each genetic component. Multiple genomic models were then tested in a predictive application testing a simple and full genetic model (GCA, and GCA+SCA, respectively) for the same tuber traits from **chapter 2** along with tuber dry matter content. Testing the predictive model performance between each model showed no additional benefit through the addition of the SCA component in the model. Additionally, the magnitude of genetic residual variance relative to the SCA variance suggested other possible higher-order genetic interactions, most notably in total tuber number and dry matter content. Despite this, genomic prediction was sufficiently powerful solely on the basis of GCA in the prediction of a hybrid's genetic value.

These prediction models were extended further in **chapter 5** by examining several other statistical paradigms and assessing any benefits to multiallelic marker information. Specific attention was taken to examining the utility of identity-by-descent (IBD) information derived from deep pedigree information linking ancestral founders to the parents of hybrids along with multiallelic identity-by-state (IBS) information. Both types of marker information were compared with conventional biallelic SNPs using traditional shrinkage-based models along with more complicated kernel prediction. The performance of SNP-based prediction models were superior over their multiallelic counterparts for all traits. Marked

differences in prediction accuracy were especially apparent between the SNP and IBD models. Similarly, little differences between the different modelling methods were found with exception that the Gaussian kernel tended to maximize prediction accuracy regardless of trait and markerset. These results would suggest that in the context of genomic prediction, simpler marker parameterizations tend to yield more consistent prediction outcomes. Despite these results, multi-allelic marker information are likely to continue to have a prominent role in inferential applications such as multi-parental population (MPP) mapping[29].

Methods of selection

Modern breeding programs are met with unbounding choice with regard to different technologies. This is especially true for molecular marker and marker-based methods of selection. Marker-assisted selection (MAS) and other marker-based techniques have revolutionized modern breeding programs by enabling precise selection of desirable traits at early stages of development. In the context of hybrid potato breeding, these methods leverage molecular markers to identify and select parental lines with favorable alleles, thereby accelerating genetic gain.

The topic of selection strategy in hybrid potato was inspected in **chapter 4**. The exercise comprised of comparing several molecular marker-assisted strategies based upon their relative efficiency which primarily depended on the prediction accuracy of the underlying model used. The primary aim was to assess whether marker-assisted selection could be as efficient as the genome-wide regression models utilized in **chapter 3**. Three strategies were considered: a marker-assisted selection strategy (`MA`), genome-wide prediction (`GW`), and a single-SNP control strategy (`PC`). Each of these were used to predict the value of an inbred parent. Based upon a forward-selection procedure, 33 unique QTL were found across three traits (with total tuber number being excluded). These QTL were able to pick up between 54 and 56 per cent of the total genetic variation and were the basis for prediction in the models used in `MA`. When examining each strategy's selection accuracy, the `GW` strategy was evidently superior for all three traits with the relative efficiency of `MA` being between 0.89 to 0.95 relative to `GW`. Because these strategies have different costs associated with them, primarily related to genotyping cost, these were also considered. When the genotyping costs were integrated into the relative efficiency, the `MA` appeared more favourable.

6.2 Wider Implications

Hybrid Breeding Schema

Two important questions broached in **chapter's 2 & 3** dealt with the nature of gene action in hybrid potato and what population and breeding strategy should be leveraged to effectively breed for complex trait improvement. Both of these questions impact the future of what *kind* of hybrid crop potato might be. While the scope of this thesis is limited by the relatively narrow genetic background sampled to initialize these inbred

populations[30], these results can still be taken into consideration to inform the strategy of a hybrid breeding program.

One of the major findings **chapter 2** was a distinct lack of SCA variance found among the panel of 806 hybrids. This was further confirmed in **chapter 3** in a genomic model were the SCA variance was smaller than both the GCA and genetic residual variance components. As remarked in these chapters, this would indicate a lack of non-additive gene action at work among our panel of F1 hybrids. This could be characteristic of very little population structure among the inbreds (as confirmed by the population-based analyses in **chapter 3**), however, this explanation is not wholly satisfactory. It has been observed in other hybrid crops that SCA variance tends to be more present in a complex trait architecture where heterotic pools are not genetically distinct[31]. Contrary to this, my results follow the more general pattern where the dominance variance appears to be nothing more than a sort of *"residual"* with most of the genetic variation being allocated to the additive variance[32]. As a complementary argument, there are also many examples of higher-order genic interactions which manifest in statistical models as additive genetic variance[33,34]. This too could be contributing to what was observed earlier in both phenotypic analyses and subsequent predictive modelling. It is worth acknowledging that this whole discussion hinges on the assumption of a direct relationship of estimated statistical and genetic parameters derived therefrom[35]. It is out of scope to defend this premise of applied genetics here, but there are many examples of such genetic parameters being accurate enough to guide decision-making in breeding applications[36,37]. In other words, "*While these [assumptions] are formally unrealistic, methods work.*"[38].

Perhaps most pertinent to the discussion of hybrid breeding is with regard to heterotic pool development and the expediency of multiple 'pools' in potato. The development of pools in crops like maize or sorghum was a relatively unguided process where complementation between distinct genetic groups was first observed often with pedigree-breeding, and then further developed with more formal methods of population improvement[39]. Hybrid crops are primarily the product of a multi-pool system, however, this is dependent upon multiple conditions related both to complex trait improvement and the additional costs of logistics in multi-pool breeding. Multiple simulation studies have examined the topic of heterotic pool development and are worth consideration for potato. Simulation of multiple tetraploid and diploid breeding program designs found that a two pool strategy based around GCA-based selection was effective in an inbred-hybrid program. However, this was dependent on the trait's degree of dominance in addition to hybrid schemas being more capital intensive[40]. The authors also remarked on the sensitivity of this strategy not only to cycle length (especially if reliant on phenotypic-based screening), but also the time required to generate fully inbred pools.

The question of how pools should be generated is also open for deliberation. In terms of quantitative trait improvement, multiple *splitting* strategies have been suggested to be equivalent including even random pool assignment, at least among selfing crops[41]. A more important consideration than genetic differentiation among pools in potato will be the market segment requirements. These will strongly govern pool development around trait targets which are widely discordant. For example, it would be likely cumbersome

to produce parental lines for ware and quick service markets from the same pool based upon the diverging product profiles of each[42]. Rather than seeing heterotic pooling as some necessity, it is instead the working out of a comprehensive strategy dependent on the robust production of inbred lines, strong fertility characteristics, and a reliable cytoplasmic male sterility mechanism[43,44]. These are the core factors which are decisive to the choice of single, two, or other multi-pool breeding schemas. The topic of fertility is considered further in Section 6.3.

Multivariate applications for potato breeders

Multivariate methods have already proven to benefit animal and crop breeding by enabling simultaneous improvement of multiple traits while delineating their genetic, environmental, and residual correlations. These approaches are particularly valuable in potato breeding, where traits such as tuber yield, tuber quality parameters, and disease resistance often exhibit complex interdependencies. By integrating genetic covariances, breeders can design selection indices that maximize genetic gain across traits, thereby optimizing resource allocation and accelerating progress toward breeding objectives. While these methods have been widely adopted in other field crops, their application in potato remains under-explored.

To lend some credence to multivariate selection in potato, let us consider what multivariate selection would involve using the trait genetic covariances reported from **chapter 2** (Figure 2.5) including estimated GCA covariance matrices together with their full phenotypic variance matrix (P). Assuming the GCA variances estimated here would be roughly equivalent to those estimated from a test cross schema in a hybrid breeding program ($G \approx V_{gca}$), the selection response on inbred parents could be estimated for future inbred development using the multivariate breeder's equation as expressed by[45]. If we performed truncation selection on average tuber volume ($s_{tv} = 2\ cm^3$) with no direct selection on total tuber yield and tuber number constant ($s_{ty} = 0\ tonnes \cdot Ha^{-1}$ & $s_{tn} = 0\ tubers$), and conducted inter-mating between selected candidates, then the expected selection response (R) next test cross cycle could be estimated as:

$$R = G \cdot P^{-1} \cdot S$$

$$R = \begin{bmatrix} 11.54 & 7.27 & 4.95 \\ 7.27 & 11.13 & 67.2 \\ 4.95 & 67.2 & 629.51 \end{bmatrix} \cdot \begin{bmatrix} 0.17 & -0.22 & 0.02 \\ -0.22 & 0.34 & -0.03 \\ 0.02 & -0.03 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1.09 \\ 0.64 \\ 0.07 \end{bmatrix}$$

This would indicate an increase of 1.1 cm^3 in tuber volume next test cross cycle with a minor increase in total tuber yield and little change in tuber number. Note that not only has this increased our selection response when used over the univariate alternative (0.9 $= h^2 s$), but we are also able to evaluate the impact of indirect selection among the other traits. This is invaluable both in forecasting and breeding strategy development. More

attention is needed here to make these methods more practical to wield as well as scalable in applied settings. With regard to practicality, selection indices are one of the primary tools used to reduce a breeder's multi-dimensional trait space into a singular index with a variety of methods proposed[46,47]. Speaking to implementability, the most challenging step in this process is the estimation of the trait genetic covariance. This tends to become quite unwieldy using traditional linear mixed modeling methods in higher-dimensional spaces making other estimation methodology more attractive[48,49].

This same exercise has utility in other facets of potato breeding. One recurrent hurdle in potato evaluation is the lack of genetic correlation between the initial seedling stages and subsequent clonal generations[50]. These differences are not restricted to tubers from these generations but are present earlier with major differences in plant architecture suggesting altered sink-source dynamics between generations[51]. This is true both for phenotypic evaluation and genetic parameter estimation and all this significantly constrains early selection efficiency within potato breeding programs[8,52]. Multivariate estimation of a candidate's *genetic value* (relevant in clonal programs) or *breeding value* (relevant for both clonal and hybrid breeding schemas) jointly across multiple generations could be a valuable extension of multivariate selection. The estimation of these inter-generational covariance structures together with traditional multivariate methods could potentially augment forecasting the potential of a varietal candidate earlier in selection cycles.

Successful genomic prediction in hybrid potato

Genomic prediction has already accelerated genetic progress in multiple hybrid crops[53]. **Chapters' 3, 4, and 5** have several implications for genomic prediction applications in potato as a hybrid crop. Considering factors related to parameterization of genetic model or type of modelling framework (what I will call more generally, *model choice*), it is not the *decisive* factor in the outcome of genomic prediction strategy. Having reviewed extensions of the traditional GBLUP, various shrinkage-based estimation methods, and one implementation of the Gaussian kernel, similar performance was observed with only marginal improvement between models. This is in keeping with many similar studies in other row crops as well as observed in tetraploid potato[54–58]. Whether the model is an extension of GBLUP or a new member of the Bayesian alphabet, there is no clear or distinct advantage observed for most traits.

As documented in other crops, often more important than the specific genetic parameterisations of a model is the actual composition of its training set[59]. In many breeding programs, training sets often arise from the breeding material itself utilizing previous breeding cycles to estimate the population genetic covariance or marker effects for a genomic model. Within hybrid breeding schemas, training sets are frequently developed from test cross blocks which often have some factorial structure (e.g. **t** testers by **x** new candidates) or sparse modification[60]. These are used both for the selection of novel lines within pools as well as for the prediction of hybrid crosses. **Chapters' 3, 4, and 5** all utilized four field trials which utilized just under 800 F1 hybrids which were the progeny of 456 inbred parents in a sparse mating design. While it could be demonstrated in an earlier chapter that the genetic models based upon these training sets were technically

identifiable[61], the sparsity of the crossing block likely led to bias in the estimation of our genetic variances and of any predictions of the genetic effects[62]. Despite this, our training set still enabled genomic prediction of hybrid performance and can likely be improved with further optimization. Aside from incorporating future cycles to augment the training set, there are many other methods relevant to increasing the phenotypic and genetic variance in a training set[63–65].

Related to training set composition is the topic of molecular marker information density. These questions are often pursued along the lines of predictive benefits in genomic modelling, the hope being that all genetic variation can be interrogated by some ideal probeset. While this cannot be fully achieved[66], molecular marker composition is a worthwhile question in efficient selection applications. In **Chapter 4**, I considered a procedure for comparing classical marker-assisted selection over genomic prediction for several quantitative traits. One of the primary conclusions of this research was that there are multiple complex traits in potato which could be effectively selected for with only a handful of molecular markers. This schema was relatively simple and could be augmented through a more robust simulation of costs, genotype-by-environment effects, and trial design[67]. A point should also be made regarding our higher marker densities, that being, these were still quite low. Many predictive modelling studies have used marker densities in the magnitude of hundreds of thousands or million of molecular markers[68,69]. Having said this, even tetraploid studies have shown that small marker panels are still capable of yielding acceptable results in genomic prediction and other trait discovery applications which is in keeping with what I found in my experimental chapters[70,71].

Aside from marker density, molecular marker content has captured the attention of many geneticists. Whether it be the inclusion of functional genetic information, enrichment of major QTL, or multi-omic data, all seek to include biologically relevant information to bolster genetic applications[72–74]. In this vein, **Chapter 5** attempted genomic prediction based upon multiallelic predictors through the inclusion of haplotag and IBD profiles. Strategies for the development of multiallelic probe design and genetic models have become a topic of interest in potato as an answer to potato's genomic diversity[75]. While recent pan-genome studies have shown less haplotype diversity among released potato varieties (particularly among European cultivars), potential applications for multiallelic molecular data are still to be identified[76]. Statistical models developed with the haplotag probeset showed nearly identical performance with models using traditional SNP-based predictors. These results would suggest little benefit in the inclusion of multiallelic predictors for genome-wide prediction based upon my survey of models used here. The performance of the IBD models were notably lower than expectation, but as noted already, this could be the product of high uncertainty for a subset of ancestral founders. Whether or not IBD information can extend genomic prediction applications, it is worthwhile to reiterate that founder composition in a breeding program can be utilized in pre-breeding efforts, understanding selection outcomes, and genetic diversity management.

6.3 Future research

There are many facets of potato breeding that have not been sufficiently addressed by current quantitative genetic frameworks. Here future applications of quantitative genetics are considered with particular emphasis on their role in improved fertility, pre-breeding methodology, and breeding risk assessment for more robust potato breeding.

Fertility Mechanisms

Fertility is a crucial factor for any seed-based crop. The success of hybrid breeding systems are particularly dependent on seed cost price which is chiefly the product of affordable and reliable seed production systems[77,78]. Most research in potato fertility up until this time has *rightly* been focussed on large effect loci like *Sli* or important genes in wider fertility modules like *StCDF1*[79–81]. As these loci continue to be used and fixed in breeding populations, an important future step will be to survey a broader array of relevant traits in diploid potato and assess them beyond QTL[82]. The typical targets for fertility in hybrid crops are pollen shed, sufficient pollen viability, and synced male and female flower opening, such that seed production is unencumbered[13,83]. In conjunction with these traits, a viable cytoplasmic male sterility (CMS) system is also crucial for F1 hybrid seed production. Not only does this aid in circumventing female line emasculation costs & minimizing F1 seed contamination, but also reduces berry set in the F1 field crop thus eliminating seed-borne volunteering post harvest. Recently, a CMS system has been proposed for hybrid potato using the *antherless* locus, *Al*, on chromosome 6 with male sterility expressed in germplasm containing the P cytoplasm[44]. Hybrids with P cytoplasm and homozygous at the *antherless* with *alal* showed significantly lower rates of berry set and less seed yields relative to fertile hybrids. One operational concern with such a strategy is that it requires not only a single locus to be fixed in both maternal and paternal lines, but all maternal lines must also contain the P cytoplasm. It should also be noted that while there is not a clear elucidation of cytoplasm's exact role on agronomic and economic traits, there is evidence of cytoplasm affecting maturity, starch content, and late blight[84]. This would have serious ramifications if transferring a P cytoplasm to a breeding program's elite maternal lines has an associated *plastid-drag* synonymous to the use of QTLs contained in large introgression regions. Further validation of the CMS model can help assess cytoplasm substitution's risk on other economic traits in hybrid potato.

Pre-breeding for Quantitative Improvement

The role of pre-breeding is often neglected with regard to complex trait improvement of potato. Pre-breeding strategies in diploid potato are particularly challenging as you are frequently interested in novel material with a higher ploidy. Pre-breeding has at least three problems: (1) How new germplasm should be screened, (2) How should ploidy reduction be done, and (3) How should that material be incorporated into a breeding program. None of these questions are very straightforward with regard to a diploid potato breeding program and is decorated by multiple hazards. In a more familiar crop breeding scenario, such as maize, if a new program is being initiated, evaluating a base population of lan-

draces before introgressing into your breeding program is a sensible strategy for trait improvement[85]. However, in diploid populations, novel tetraploid germplasm must first be subjected to androgenesis or *prickle pollination* before dihaploids can be introduced into a program[86,87]. This process not only disrupts the original trait architecture of that original tetraploid donor, but these dihaploids frequently bear little to no resemblance to the original donor making their original evaluation dubious.

Multiple screening approaches have been suggested, all of which, depend on molecular marker information of some form. One approach proposed in banana and potato suggests an extension of GBLUP whereby multi-ploidy training sets could enable prediction from the 4x to 2x ploidy levels[88,89]. The prediction accuracy observed in these studies appear modest, but whether the endeavour of dedicated tetraploid field evaluation for an augmented across-ploidy training set is altogether justified is unclear. To avoid this, the IBD estimation methodology from **chapter 5** could be considered as a potential methodology for *linking* the genetic value of dihaploids to their tetraploid forbearer's. IBD estimation of deep pedigrees in diploids has been realized in both simulated populations as well as real[90]. Additionally, IBD estimation has been conducted in shallow structured populations in tetraploids[91,92]. Very little methodological development would be needed to extend the IBD tracing from tetraploid to related dihaploids enabling the estimation of allelic effects (in diploid populations) from specific tetraploid haplotypes. If a novel donor has already been introgressed into a program, IBD profiles could be leveraged and even extended to related tetraploids.

Both these methods essentially seek to explain the performance of diploids by some molecular proxy in the tetraploid either through a genetic covariance structure or via explicit IBD probabilities. One last consideration for future improvements in pre-breeding is captured in what is often called the next stage of breeding or *"Breeding 4.0"*[93]. The central premise is that a crop's *ideotype* can be designed from the ground up based upon a library of known functionally annotated haplotypes which are then combined into an optimal cultivar via cis or transgenic technologies[94]. Breeding is no longer a practice of finding the proverbial needle in the haystack, but is instead the direct forging of a variety from constituent components. This would make the role of pre-breeding primarily focused on identifying those components. Recent endeavours in diploid potato breeding could be described as examples of such crop design, most notably,[95]. Populations are screened on the basis of functional genetic load based upon some algorithmic method (e.g. SIFT), and are then used to hasten the process of inbred production. Whether this approach can produce a sufficient volume of inbreds to initialize a breeding program is to be seen, but it is a tool which is already being used with the intent to design future potato cultivars.

Formalization of Risk

When considering a novel technology or more efficient breeding schema, the first question often asked is *"What kind of genetic gains can be expected with this implementation"*? This is sensible but without asking *"what are the risks of such an implementation"* biometricians run the risk of optimistic forecasting. This addendum is especially pertinent for potato breeders due to many of the aforementioned challenges, chief among them, envi-

ronmental sensitivity[96]. Formalizing the estimation of explicit risk of a breeding strategy is the province of quantitative genetics and thus deserves greater attention. For example, decreasing the cycle length of a schema by reducing the number of years of trial evaluation is a classic route used for increasing genetic gain. However, this also stands to increase the uncertainty around a candidate if they have been exposed to an insufficient number of environments. Embracing other tools might also bring risk to the forefront. Factor analytic and random regression models have been lauded for their ability to assess genotype stability and adaptability, and offer incredible potential for evaluation in potato[97,98]. Looking to methods utilized outside of plant breeding such as the use of hazard models could also be useful for assessing the long-term stability of a potential potato candidate[99].

6.4 Conclusion

I present here a first examination into the quantitative nature of several complex traits in hybrid potato. Additionally, through the survey of a variety of statistical models, I offer a framework for evaluating hybrid potato genetic architecture, genotype-by-environment interaction, and genetic progress. This thesis has also demonstrated the transformative potential of biometrical methods in advancing hybrid potato breeding. Our findings underscore the importance of building on traditional statistical formulations for both inferential and predictive applications in potato breeding. These insights not only bridge critical gaps in potato breeding but also align with broader trends in crop improvement, where quantitative genetics and genomic technologies are reshaping traditional paradigms. Moving forward, the integration of these methods into breeding programs promises to enhance precision, reduce cycle times, and unlock the full potential of hybrid potato as a sustainable crop for future agriculture.

References

1. Bradshaw JE. Review and Analysis of Limitations in Ways to Improve Conventional Potato Breeding. Potato Research [Internet] 2017 [cited 2021 Jun 28];60:171–93. Available from: http://www.tsl.ac.uk/news/new-potato-at-the-

2. Adams JR, de Vries ME, Zheng C, van Eeuwijk FA. Little heterosis found in diploid hybrid potato: The genetic underpinnings of a new hybrid crop. G3: Genes, Genomes, Genetics [Internet] 2022;12(6). Available from: https://www.ncbi.nlm.nih.gov/pubmed/35460241

3. Adams J, De Vries M, Van Eeuwijk F. Efficient Genomic Prediction of Yield and Dry Matter in Hybrid Potato. Plants [Internet] 2023 [cited 2023 Nov 9];12(14):2617. Available from: https://www.mdpi.com/2223-7747/12/14/2617

4. Schoemaker DL, Lima DC, de Leon N, Kaeppler SM. Modeling the impact of resource allocation decisions on genomic prediction using maize multi-environment data. Crop Science [Internet] 2024 [cited 2025 Sep 6];64(5):2748–67. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/csc2.21305

5. Gopal J. Challenges and Way-forward in Selection of Superior Parents, Crosses and Clones in Potato Breeding. Potato Research [Internet] 2015 [cited 2023 Jan 20];58(2):165–88. Available from: https://link-springer-com.ezproxy.library.wur.nl/article/10.1007/s11540-015-9292-6

6. Paget MF, Alspach PA, Anderson JA, D, Genet RA, Braam WF, et al. Replicate allocation to improve selection efficiency in the early stages of a potato breeding scheme. Euphytica [Internet] 2017 [cited 2024 Nov 21];213(9):1–15. Available from: https://www.proquest.com/docview/1934680784/abstract/3B29004E4B074DC0PQ/1

7. Zystro J, Colley M, Dawson J. Alternative Experimental Designs for Plant Breeding [Internet]. In: Plant Breeding Reviews. Wiley; 2018 [cited 2020 Jan 4]. page 87–117.Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119521358.ch3

8. Davies HT, Johnston GR. Reliability of potato selection in the first clonal generation. American Potato Journal [Internet] 1974 [cited 2024 Nov 21];51(1):8–11. Available from: https://doi.org/10.1007/BF02852022

9. Pallais N. True Potato Seed: Changing Potato Propagation from Vegetative to Sexual. HortScience 1991;26(3):239–41.

10. Storck L, Lopes SJ, Lúcio AD, Cargnelutti Filho A. Optimum plot size and number of replications related to selective precision. Cienc Rural [Internet] 2011 [cited 2025 Sep 6];41:390–6. Available from: https://www.scielo.br/j/cr/a/3T8B3TxTLGQXZqPdssPKHRD/?format=html&lang=en

11. Stockem JE, Korontzis G, Wilson SE, Vries ME de, Eeuwijk FA, Struik PC. Optimal plot dimensions for performance testing of hybrid potato in the field. Potato Research 2021;

12. Anderson JaD, Howard HW. Effectiveness of selection in the early stages of potato breeding programmes. Potato Res [Internet] 1981 [cited 2025 May 27];24(3, 3):289–99. Available from: https://link-springer-com.ezproxy.library.wur.nl/article/10.1007/BF02360366

13. Ortiz R. Breeding for potato production from true seed. 1997 [cited 2024 Nov 20];Available from: https://www.cabidigitallibrary.org/doi/full/10.5555/19971610712

14. Gu J, Struik PC, Evers JB, Lertngim N, Lin R, Driever SM. Quantifying differences in plant architectural development between hybrid potato (Solanum tuberosum) plants grown from two types of propagules. Annals of Botany [Internet] 2024 [cited 2025 Nov 2];133(2):365–78. Available from: https://academic.oup.com/aob/article-abstract/133/2/365/7473623

15. van Dijk LCM, Lommen WJM, de Vries ME, Kacheyo OC, Struik PC. Hilling of Transplanted Seedlings from Novel Hybrid True Potato Seeds Does Not Enhance Tuber Yield but Can Affect Tuber Size Distribution. Potato Res [Internet] 2021 [cited

2025 May 28];64(3, 3):353–74. Available from: https://link.springer.com/article/10.1007/s11540-020-09481-x

16. Gilmour AR, Cullis BR, Verbyla AP. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. Journal of Agricultural, Biological, and Environmental Statistics [Internet] 1997 [cited 2025 May 28];2(3):269–93. Available from: https://www.jstor.org/stable/1400446

17. Velazco JG, Rodríguez-Álvarez MX, Boer MP, Jordan DR, Eilers PHC, Malosetti M, et al. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. Theoretical and Applied Genetics [Internet] 2017 [cited 2021 Jun 8];130(7):1375–92. Available from: https://cran.r-project.org/

18. Piepho HP, Möhring J, Schulz-Streeck T, Ogutu JO. A stage-wise approach for the analysis of multi-environment trials. Biometrical Journal [Internet] 2012 [cited 2025 May 28];54(6):844–60. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201100219

19. Gogel B, Smith A, Cullis B. Comparison of a one- and two-stage mixed model analysis of Australia's National Variety Trial Southern Region wheat data. Euphytica [Internet] 2018 [cited 2025 May 28];214(2, 2):1–21. Available from: https://link.springer.com/article/10.1007/s10681-018-2116-4

20. Damesa TM, Möhring J, Worku M, Piepho HP. One Step at a Time: Stage-Wise Analysis of a Series of Experiments. Agronomy Journal [Internet] 2017 [cited 2023 Feb 17];109(3):845–57. Available from: https://acsess-onlinelibrary-wiley-com.ezproxy.library.wur.nl/doi/10.2134/agronj2016.07.0395

21. Bertan I, de Carvalho FI, Oliveira AC de. Parental selection strategies in plant breeding programs. Journal of crop science and biotechnology [Internet] 2007 [cited 2025 Jun 11];10(4):211–22. Available from: https://koreascience.kr/article/JAKO200716637994242.page

22. Neele AEF, Nab HJ, Louwes KM. Identification of superior parents in a potato breeding programme. Theoretical and Applied Genetics [Internet] 1991 [cited 2021 Oct 9];82(3):264–72. Available from: https://link.springer.com/article/10.1007/BF02190611

23. Bradshaw JE, Dale · MFB, Mackay · GR. https://doi.org/10.1007/s00122-003-1219-yUse of mid-parent values and progeny tests to increase the efficiency of potato breeding for combined processing quality and disease and pest resistance. Theor Appl Genet 2003;107:36–42.

24. Plaisted RL, Sanford L, Federer WT, Kehr AE, Peterson LC. https://doi.org/10.1007/BF02871402Specific and general combining ability for yield in potatoes. American Potato Journal 1962;39(5):185–97.

25. Tai GCC. ESTIMATION OF GENERAL AND SPECIFIC COMBINING ABILITIES IN POTATO. Can J Genet Cytol [Internet] 1976 [cited 2025 May 17];18(3):463–70. Available from: http://www.nrcresearchpress.com/doi/10.1139/g76-056

26. Killick RJ. Genetic analysis of several traits in potatoes by means of a diallel cross. Annals of Applied Biology [Internet] 1977 [cited 2021 Oct 9];86(2):279–89. Available from: https://onlinelibrary-wiley-com.ezproxy.library.wur.nl/doi/full/10.1111/j.1744-7348.1977.tb01841.x

27. Gebhardt C. Bridging the gap between genome analysis and precision breeding in potato. Trends in Genetics [Internet] 2013 [cited 2023 Feb 1];29(4):248–56. Available from: https://www.sciencedirect.com/science/article/pii/S0168952512001904

28. Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C, et al. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity [Internet] 2017 [cited 2025 May 28];118(3):302–10. Available from: https://www.nature.com/articles/hdy201687

29. Li W, Boer MP, Zheng C, Joosen RVL, van Eeuwijk FA. An IBD-based mixed model approach for QTL mapping in multiparental populations. Theoretical and Applied Genetics 2021 134:11 [Internet] 2021 [cited 2021 Nov 1];134(11):3643–60. Available from: https://link.springer.com/article/10.1007/s00122-021-03919-7

30. Lindhout P, de Vries M, ter Maat M, Ying S, Viquez-Zamora Marcela, van Heusden S. https://doi.org/10.1201/9781351114455-10/HYBRID-POTATO-BREEDING-IMPROVED-VARIETIES-GEFU-WANG-PRUSKIHybrid potato breeding for improved varieties. In: Wang-Pruski Gefu, editor. Achieving sustainable cultivation of potatoes Volume 1. Burleigh Dodds Science Publishing; 2018. page 119–42.

31. Zhao Y, Mette MF, Reif JC. Genomic selection in hybrid breeding. Plant Breeding [Internet] 2015 [cited 2023 May 1];134(1):1–10. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/pbr.12231

32. Bernardo R. Bandwagons I, too, have known. Theoretical and Applied Genetics [Internet] 2016;129(12):2323–32. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27681088

33. Hill WG, Goddard ME, Visscher PM. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. PLOS Genetics [Internet] 2008 [cited 2025 Oct 2];4(2):e1000008. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000008

34. Monnahan PJ, Kelly JK. Epistasis Is a Major Determinant of the Additive Genetic Variance in Mimulus guttatus. PLOS Genetics [Internet] 2015 [cited 2025 Oct 2];11(5):e1005201. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005201

35. Campos G de los, Sorensen D, Gianola D. Genomic Heritability: What Is It? PLOS Genetics [Internet] 2015 [cited 2025 Sep 27];11(5):e1005048. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005048

36. Walsh B. Population-and quantitative-genetic models of selection limits. Plant breeding reviews [Internet] 2004 [cited 2025 Oct 2];24(1):177–226. Available from: https://

scholar.archive.org/work/galt5pvqqvdvvaku2o7mrtuf74/access/wayback/http://nitro.biosci.arizona.edu:80/workshops/Aarhus2006/pdfs/WalshLT.pdf

37. Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM. Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. Genetics [Internet] 2013 [cited 2025 Oct 2];193(2):347–65. Available from: https://doi.org/10.1534/genetics.112.147983

38. Hill WG. Understanding and using quantitative genetic variation. Philosophical Transactions of the Royal Society B: Biological Sciences [Internet] 2010 [cited 2025 Oct 2];365(1537):73–85. Available from: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2009.0203

39. Duvick DN, Smith JSC, Cooper M. Long-term Selection in a Commercial Hybrid Maize Breeding Program. Plant Breeding Reviews 2004;24(2):109–52.

40. Labroo MR, Endelman JB, Gemenet DC, Werner CR, Gaynor RC, Covarrubias-Pazaran GE. Clonal diploid and autopolyploid breeding strategies to harness heterosis: Insights from stochastic simulation. Theoretical and Applied Genetics [Internet] 2023 [cited 2023 Jun 9];136(7):1–19. Available from: https://link.springer.com/article/10.1007/s00122-023-04377-z

41. Cowling WA, Gaynor RC, Antolín R, Gorjanc G, Edwards SM, Powell O, et al. In silico simulation of future hybrid performance to evaluate heterotic pool formation in a self-pollinating crop. Scientific Reports [Internet] 2020 [cited 2023 May 6];10(1):4037. Available from: https://www.nature.com/articles/s41598-020-61031-0

42. Keijbets MJH. Potato Processing for the Consumer: Developments and Future Challenges. Potato Res [Internet] 2008 [cited 2025 Aug 9];51(3):271–81. Available from: https://doi.org/10.1007/s11540-008-9104-3

43. McGrath JM, Panella L. Sugar Beet Breeding [Internet]. In: Goldman I, editor. Plant Breeding Reviews. Wiley; 2018 [cited 2024 Mar 20]. page 167–218.Available from: https://onlinelibrary.wiley.com/doi/10.1002/9781119521358.ch5

44. Eggers EJ, Su Y, van der Poel E, Flipsen M, de Vries ME, Bachem CWB, et al. Identification, Elucidation and Deployment of a Cytoplasmic Male Sterility System for Hybrid Potato. Biology [Internet] 2024 [cited 2025 Oct 25];13(6):447. Available from: https://www.mdpi.com/2079-7737/13/6/447

45. Lande R, Arnold SJ. The Measurement of Selection on Correlated Characters. Evolution [Internet] 1983 [cited 2025 May 30];37(6):1210–26. Available from: https://www.jstor.org/stable/2408842

46. Kempthorne O, Nordskog AW. Restricted Selection Indices. Biometrics [Internet] 1959 [cited 2025 May 30];15(1):10–9. Available from: https://www.jstor.org/stable/2527598

47. Bulmer MG. Selection Indeces. In: The mathematical theory of quantitative genetics. Oxford, UK: Clarendon Press; 1981. page 255.

48. Blows M, Walsh B. Spherical Cows Grazing in Flatland: Constraints to Selection and Adaptation. In: van der Werf J, Graser HU, Frankham R, Gondro C, editors. Adaptation and Fitness in Animal Populations. Dordrecht: Springer; 2009. page 83–101.

49. Runcie DE, Mukherjee S. Dissecting High-Dimensional Phenotypes with Bayesian Sparse Factor Analysis of Genetic Covariance Matrices. Genetics [Internet] 2013 [cited 2025 May 29];194(3):753–67. Available from: https://doi.org/10.1534/genetics.113.151217

50. Maris B. Correlations within and between characters between and within generations as a measure for the early generation selection in potato breeding. Euphytica [Internet] 1988 [cited 2024 Nov 21];37(3):205–24. Available from: https://doi.org/10.1007/BF00015117

51. Gu J, Evers JB, Struik PC, Driever SM. Architectural Differences between Hybrid Potato Plants Grown from True Seeds and Seedling Tubers Do Not Influence Leaf Photosynthetic Traits. Potato Res [Internet] 2025 [cited 2025 Oct 25];Available from: https://doi.org/10.1007/s11540-025-09930-5

52. Gopal J. General combining ability and its repeatability in early generations of potato breeding programmes. Potato Research [Internet] 1998 [cited 2024 Nov 21];41(1):21–8. Available from: https://www.proquest.com/docview/754151133/abstract/AD7024E9634C42D0PQ/1

53. Labroo MR, Studer AJ, Rutkoski JE. https://doi.org/10.3389/FGENE.2021.643761Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. Frontiers in Genetics 2021;12:234.

54. Sverrisdóttir E, Byrne S, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, et al. Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. Theor Appl Genet [Internet] 2017 [cited 2025 Jul 25];130(10):2091–108. Available from: https://doi.org/10.1007/s00122-017-2944-y

55. Endelman JB, Carley CAS, Bethke PC, Coombs JJ, Clough ME, da Silva WL, et al. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics [Internet] 2018 [cited 2022 Nov 14];209(1):77–87. Available from: https://doi.org/10.1534/genetics.118.300685

56. Amadeu RR, Ferrão LFV, Oliveira I de B, Benevenuto J, Endelman JB, Munoz PR. https://doi.org/10.2135/cropsci2019.02.0138Impact of Dominance Effects on Autotetraploid Genomic Prediction. Crop Science 2019;0(0):1–9.

57. Wilson S, Zheng C, Maliepaard C, Mulder HA, Visser RGF, van der Burgt A, et al. Understanding the Effectiveness of Genomic Prediction in Tetraploid Potato. Frontiers in Plant Science [Internet] 2021 [cited 2022 Nov 15];12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.672417

58. Schrauf MF, de los Campos G, Munilla S. Comparing Genomic Prediction Models by Means of Cross Validation. Front Plant Sci [Internet] 2021 [cited 2025 Aug 16];12.

Available from: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.734512/full

59. Dias KOG, Piepho · HP, Guimarães · LJM, Guimarães · PEO, Parentoni · SN, Pinto · MO, et al. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theoretical and Applied Genetics [Internet] 2019 [cited 2019 Dec 4];1:3. Available from: https://doi.org/10.1007/s00122-019-03475-1

60. Jarquin D, Howard R, Crossa J, Beyene Y, Gowda M, Martini JWR, et al. Genomic Prediction Enhanced Sparse Testing for Multi-environment Trials. G3 Genes|Genomes|Genetics [Internet] 2020 [cited 2025 Jul 25];10(8):2725–39. Available from: https://doi.org/10.1534/g3.120.401349

61. Xenakis J. Special Topics For Recombinant Inbred Intercross Data: Model Identifiability, Hypothesis Testing and Compositional Methods [Internet]. 2019 [cited 2021 Jul 6];Available from: https://search.proquest.com/openview/28f4f585a836499e5a03e93f37b54d04/1?pq-origsite=gscholar&cbl=18750&diss=y

62. Möhring J, Melchinger AE, Piepho HP. REML-Based Diallel Analysis. Crop Science [Internet] 2011 [cited 2021 Feb 3];51(2):470–8. Available from: http://doi.wiley.com/10.2135/cropsci2010.05.0272

63. Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME. Training set optimization under population structure in genomic selection. Theor Appl Genet [Internet] 2015 [cited 2025 Aug 16];128(1):145–58. Available from: https://doi.org/10.1007/s00122-014-2418-4

64. Berro I, Lado B, Nalin RS, Quincke M, Gutiérrez L. https://doi.org/10.3835/plantgenome2019.04.0028Training population optimization for genomic selection. Plant Genome 2019;12(3).

65. Ou JH, Liao CT. Training set determination for genomic selection. Theor Appl Genet [Internet] 2019 [cited 2025 Aug 16];132(10):2781–92. Available from: https://doi.org/10.1007/s00122-019-03387-0

66. Brachi B, Morris GP, Borevitz JO. Genome-wide association studies in plants: The missing heritability is in the field. Genome Biol [Internet] 2011 [cited 2025 Aug 16];12(10):232. Available from: https://doi.org/10.1186/gb-2011-12-10-232

67. Peixoto MA, Coelho IF, Leach KA, Bhering LL, Resende Jr. MFR. Simulation-based decision-making and implementation of tools in hybrid crop breeding pipelines. Crop Science [Internet] 2024 [cited 2025 Aug 16];64(1):110–25. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/csc2.21139

68. Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. Heredity [Internet] 2014 [cited 2025 Aug 16];112(1):39–47. Available from: https://www.nature.com/articles/hdy201313

69. Weber SE, Roscher-Ehrig L, Kox T, Abbadi A, Stahl A, Snowdon RJ. Genomic prediction in Brassica napus: Evaluating the benefit of imputed whole-genome sequencing data. Genome [Internet] 2024 [cited 2025 Aug 16];67(7):210–22. Available from: https://cdnsciencepub.com/doi/full/10.1139/gen-2023-0126

70. Aalborg T, Sverrisdóttir E, Kristensen HT, Nielsen KL. The effect of marker types and density on genomic prediction and GWAS of key performance traits in tetraploid potato. Front Plant Sci [Internet] 2024 [cited 2025 Aug 16];15. Available from: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1340189/full

71. Leyva-Pérez M de la O, Vexler L, Byrne S, Clot CR, Meade F, Griffin D, et al. PotatoMASH—A Low Cost, Genome-Scanning Marker System for Use in Potato Genomics and Genetics Applications. Agronomy [Internet] 2022 [cited 2024 Sep 3];12(10):2461. Available from: https://www.mdpi.com/2073-4395/12/10/2461

72. Guo Z, Magwire MM, Basten CJ, Xu Z, Wang D. Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. Theor Appl Genet [Internet] 2016 [cited 2025 Aug 16];129(12):2413–27. Available from: https://doi.org/10.1007/s00122-016-2780-5

73. MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics [Internet] 2016 [cited 2025 Aug 16];17(1):144. Available from: https://doi.org/10.1186/s12864-016-2443-6

74. Arouisse B, Thoen MPM, Kruijer W, Kunst JF, Jongsma MA, Keurentjes JJB, et al. Bivariate GWA mapping reveals associations between aliphatic glucosinolates and plant responses to thrips and heat stress. The Plant Journal [Internet] 2024 [cited 2025 Aug 16];120(2):674–86. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.17009

75. Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, et al. Genome evolution and diversity of wild and cultivated potatoes. Nature [Internet] 2022 [cited 2025 Jan 3];606(7914):535–41. Available from: https://www.nature.com/articles/s41586-022-04822-x

76. Sun H, Tusso S, Dent CI, Goel M, Wijfjes RY, Baus LC, et al. The phased pan-genome of tetraploid European potato. Nature [Internet] 2025 [cited 2025 Aug 16];642(8067):389–97. Available from: https://www.nature.com/articles/s41586-025-08843-0

77. Mao CX, Virmani SS, Kumar I. Technological innovations to lower the costs of hybrid rice seed production. In: Siddiq EA, Muralidharan K, editors. Advances in hybrid rice technology: Proceedings of the 3rd prebreeding International Symposium on Hybrid Rice, 14-16 November 1996, Hyderabad, India. Los Baños, Philippines: International Rice Research Institute; 1998.

78. Longin CFH, Reif JC, Würschum T. Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory. Theor Appl Genet [Internet] 2014 [cited 2025 May 31];127(7, 7):1635–41. Available from: https://link.springer.

com/article/10.1007/s00122-014-2325-8

79. Clot CR, Polzer C, Prodhomme C, Schuit C, Engelen CJM, Hutten RCB, et al. The origin and widespread occurrence of Sli-based self-compatibility in potato. Theor Appl Genet [Internet] 2020 [cited 2025 May 31];133(9, 9):2713–28. Available from: https://link.springer.com/article/10.1007/s00122-020-03627-8

80. Eggers EJ, van der Burgt A, van Heusden SAW, de Vries ME, Visser RGF, Bachem CWB, et al. Neofunctionalisation of the Sli gene leads to self-compatibility and facilitates precision breeding in potato. Nature Communications [Internet] 2021 [cited 2023 Jan 20];12(1):4141. Available from: https://www.nature.com/articles/s41467-021-24267-6

81. Song L, Endelman J. https://doi.org/10.1101/2022.11.09.515871Using Haplotype and QTL Analysis to Fix Favorable Alleles in Diploid Potato Breeding. 2022.

82. Kempe K, Gils M. Pollination control technologies for hybrid breeding. Mol Breeding [Internet] 2011 [cited 2025 May 31];27(4, 4):417–37. Available from: https://link.springer.com/article/10.1007/s11032-011-9555-0

83. Longin CFH, Mühleisen J, Maurer HP, Zhang H, Gowda M, Reif JC. Hybrid breeding in autogamous cereals. Theoretical and Applied Genetics [Internet] 2012 [cited 2024 Apr 30];125(6):1087–96. Available from: http://link.springer.com/10.1007/s00122-012-1967-7

84. Sanetomo R, Gebhardt C. Cytoplasmic genome types of European potatoes and their effects on complex agronomic traits. BMC Plant Biol [Internet] 2015 [cited 2025 Nov 2];15(1):162. Available from: https://doi.org/10.1186/s12870-015-0545-y

85. Gorjanc G, Jenko J, Hearne SJ, Hickey JM. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. BMC Genomics [Internet] 2016 [cited 2023 Oct 31];17(1):30. Available from: http://www.biomedcentral.com/1471-2164/17/30

86. Uijtewaal BA, Huigen DJ, Hermsen JGTh. Production of potato monohaploids (2n=x=12) through prickle pollination. Theoret Appl Genetics [Internet] 1987 [cited 2025 Aug 2];73(5):751–8. Available from: https://doi.org/10.1007/BF00260786

87. Tai GCC, Xiong XY. Haploid production of potatoes by anther culture [Internet]. In: Maluszynski M, Kasha KJ, Forster BP, Szarejko I, editors. Doubled Haploid Production in Crop Plants: A Manual. Dordrecht: Springer Netherlands; 2003 [cited 2025 Aug 2]. page 229–34.Available from: https://doi.org/10.1007/978-94-017-1293-4_34

88. Nyine M, Uwimana B, Blavet N, Hřibová E, Vanrespaille H, Batte M, et al. Genomic Prediction in a Multiploid Crop: Genotype by Environment Interaction and Allele Dosage Effects on Predictive Ability in Banana. The Plant Genome [Internet] 2018 [cited 2025 Sep 13];11(2):170090. Available from: https://onlinelibrary.wiley.com/doi/abs/10.3835/plantgenome2017.10.0090

89. Wilson S, Malosetti M, Maliepaard C, Mulder HA, Visser RGF, van Eeuwijk F. Training Set Construction for Genomic Prediction in Auto-Tetraploids: An Example in Potato. Frontiers in Plant Science [Internet] 2021 [cited 2022 Nov 15];12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.771075

90. Li W, Boer MP, Zheng C, Joosen RVL, van Eeuwijk FA. An IBD-based mixed model approach for QTL mapping in multiparental populations. Theoretical and Applied Genetics 2021 134:11 [Internet] 2021 [cited 2021 Nov 1];134(11):3643–60. Available from: https://link.springer.com/article/10.1007/s00122-021-03919-7

91. Amadeu RR, Muñoz PR, Zheng C, Endelman JB. QTL mapping in outbred tetraploid (and diploid) diallel populations. Genetics [Internet] 2021 [cited 2023 Aug 22];219(3):iyab124. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8570786/

92. Song L, Endelman JB. Using haplotype and QTL analysis to fix favorable alleles in diploid potato breeding. The Plant Genome [Internet] 2023 [cited 2025 Sep 13];16(2):e20339. Available from: https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.20339

93. Wallace JG, Rodgers-Melnick E, Buckler ES. On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. Annual Review of Genetics [Internet] 2018 [cited 2025 May 31];52:421–44. Available from: https://www.annualreviews.org/content/journals/10.1146/annurev-genet-120116-024846

94. Varshney RK, Bohra A, Yu J, Graner A, Zhang Q, Sorrells ME. Designing Future Crops: Genomics-Assisted Breeding Comes of Age. Trends in Plant Science [Internet] 2021 [cited 2025 Sep 13];26(6):631–49. Available from: https://www.cell.com/trends/plant-science/abstract/S1360-1385(21)00069-8

95. Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, et al. https://doi.org/10.1016/J.CELL.2021.06.006Genome design of hybrid potato. Cell 2021;184(15):3873–3883.e12.

96. George TS, Taylor MA, Dodd IC, White PJ. Climate Change and Consequences for Potato Production: A Review of Tolerance to Emerging Abiotic Stress. Potato Res [Internet] 2017 [cited 2025 May 30];60(3, 3):239–68. Available from: https://link.springer.com/article/10.1007/s11540-018-9366-3

97. Souza MH de, Júnior JDP, Steckling SDM, Mencalha J, Dias F dos S, Rocha JR do AS de C, et al. Adaptability and stability analyses of plants using random regression models. PLOS ONE [Internet] 2020 [cited 2025 May 30];15(12):e0233200. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0233200

98. Smith AB, Cullis BR. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. Euphytica [Internet] 2018 [cited 2023 Feb 16];214(8):143. Available from: https://doi.org/10.1007/s10681-018-2220-5

99. Yazdi MH, Visscher PM, Ducrocq V, Thompson R. Heritability, Reliability of Genetic Evaluations and Response to Selection in Proportional Hazard Models. Journal of Dairy Science [Internet] 2002 [cited 2024 Nov 21];85(6):1563–77. Available from: https://www.sciencedirect.com/science/article/pii/S0022030202742264

Summary

Potato (*Solanum tuberosum L.*) currently stands as the world's third most important food crop after rice and wheat. Potato provides essential nutrition and calories to billions of people globally. Its tubers are prized not only for their versatility in culinary applications but also for their nutritional value, offering a rich source of carbohydrates, vitamins, minerals, and antioxidants. Beyond its role as a staple food, potato serves as a critical raw material for industrial applications including starch production, processed foods, and bio-based materials. Despite its essential role in international food security, genetic improvement of potato faces formidable challenges and is constrained by multiple biological and market-related factors. The crop's tetraploid nature, clonal propagation system, and complex genetic architecture complicate breeding efforts and negatively affect innovation. Furthermore, potato's susceptibility to numerous pests and diseases, coupled with its sensitivity to environmental stresses, necessitates continuous genetic innovation. These realities coupled with long breeding cycles characteristic of clonally propagated crops hamper the creation of novel potato varieties. Hybrid and diploid breeding systems have been proposed as a potential solution which could circumvent some of the issues facing traditional clonal breeding. Hybrid breeding systems are inherently true seed based which allow for faster propagation and multiplication rates in a breeding pipeline. True potato seed (TPS) can be leveraged in replicated early yield and multi-environmental trialling resulting in higher quality evaluations of important traits such as yield. This in turn increases the accuracy and potentially the speed of selection earlier in the breeding cycle. The use of TPS further enables the utilization of methodologies such as quantitative trait locus (QTL) mapping and genomic prediction (GP) which depend on high quality phenotypic data. The adoption of TPS transforms the traditional potato breeding schema into a data-driven program. However, this is all contingent on several genetic assumptions in how hybrid potato behaves. That includes a sufficient level of heterotic vigour, high heritabilities among complex traits, and consistent uniformity in the F1 hybrids. Characterising these factors are crucial for any implementation of hybrid potato. My thesis assesses these questions in the following form:

In **Chapter 1** I begin with a brief introduction of the evolutionary and recent cultivated history of potato. This is followed by a justification for genetic improvement in potato lay out current popular breeding methods for clonal potato. This includes an examination of specific factors of potato's biology which directly inform breeding; factors such as potato's tetraploidy, obligate clonal propagation, and deleterious mutation load. I then focus on hybrid breeding methodology and propose where it could potentially augment genetic

gain. However, to properly evaluate whether hybrid breeding is possible in potato, a proper quantitative evaluation must be made using extant material. This chapter concludes by describing this knowledge gap and the aims of this thesis.

A central component of hybrid breeding is the ability to conduct multi-environmental trials (MET) and the capacity to retrieve accurate estimates of merit (e.g. best linear unbiased predictor) with suitable heritabilities to warrant use as a selection criteria in a breeding program. **Chapter 2** examines this core premise for the first time in hybrid potato utilizing METs from an active breeding program. I and my coauthors begin with a thorough description of the trial data as it is one of the largest hybrid potato METs published to date. We showcased a two-step statistical evaluation of the MET where intra-trial effects were first estimated and fitted for all observations. This was followed by a secondary multi-trial model where the spatially-resolved yield data was used to estimate multiple genetical parameters, namely general and specific combining ability (GCA & SCA respectively). We found that modelling yield on GCA of the parents was sufficient for all yield components studied. All traits had significant SCA effects indicative of non-additive genetic effects which could not be captured by parental effects alone. We also examined the genetic architecture of these yield components via a multivariate linear mixed model methodology. This procedure allowed us to delineate these genetic correlations into additive and non-additive components. These results identified average tuber volume as a potentially useful trait for long-term yield gains in a hybrid potato program.

Because the trial data from **Chapter 2** was sufficient for estimating genetic parameters for hybrid potato, it was natural to extend the use of this trial data to other applications as well. Hybrid breeding is often improved through the use predictive methodologies such as genomic prediction which can increase selection accuracy, offer pre-selection capabilities, and shorten breeding cycles. Because of this, examining the current capacity of genomic prediction methods in potato dataset is critical for any hybrid program going forward. To this end, **Chapter 3** built on the two-step phenotypical modelling approach of **Chapter 2** through the incorporation of molecular marker information. To assess the quality of the molecular marker data, we conducted a thorough marker filtering analysis along with an examination of population structure among the parental lines. We found moderate population structure among the parental lines indicative of minor parental grouping, but no proper pool structure. We then created two genomic models mirroring the structure of the GCA and GCA+SCA models from **Chapter 2**. While the full GCA+SCA model should be selected on the basis of the likelihood ratio test and similar criteria, there was little gain in model performance through the incorporation of the SCA dominance matrix for the yield components considered.

The results of **Chapter 3** confirm that breeding values for parental lines had suitable accuracy for multiple yield components and could be predicted from a genomic GCA model. However, while genomic selection is undeniably efficient for the prediction of complex traits, high density molecular marker sampling are often the largest expense and logistical bottleneck when genomic selection is used. This makes genomic selection cost prohibitive especially in scenarios where pre-selection is done with thousands of potential

candidates. In such environments, marker-assisted selection (MAS) becomes much more interesting for large pre-selection panels. To investigate the utility of MAS for complex traits in hybrid potato, we consider a few selection scenarios in **Chapter 4**. Specifically, we examine the selection efficiency of a MAS strategy using previously identified QTL for several yield components versus a genome-wide selection schema. While we found that the MAS schema was less efficient than genome-wide selection, this was not guaranteed when cost was integrated into the selection efficiency statistic. One of the major findings of this chapter is that MAS is not only possible for complex traits in hybrid potato, but potentially favourable if a program is sensitive to genotyping costs.

Both **Chapter's 3 & 4** utilized a mid-density panel of biallelic SNPs which is standard both in genetic theory as well as for most marker-based applications in crop plants. All this said, there has been much interest in the use of multi-allelic marker information for popular genomic analyses including genomic prediction applications. This is especially true in potato where the genomic diversity is cited as a justification for incorporating multiallelic information in genomic models. **Chapter 5**, my final experimental chapter, examined the utility of multiallelic molecular marker information for genome-wide predictive models. We examined multiple forms of multiallelic marker information, namely, identity-by-state (IBS) marker haplotags and estimated founder origin identity-by-descent (IBD) variants. Using a traditional cross validation methodology, we tested both forms of multiallelic marker data against biallelic SNP information for a profile of yield components. Interestingly, there was no gain in model performance using the haplotag markerset with a clear decrease in performance using the founder IBDs as model predictors. We outline a few potential reasons for these results and offer alternative applications outside the prediction of GEBVs. A major finding was that multiallelic marker information have little merit in the genomic prediction of yield in potato.

I conclude this thesis with a general discussion and retrospective in **Chapter 6**. I take the time to examine the implications of these results for future hybrid potato breeding and locate questions which are still unaddressed at this point in time.



About the author

James was born in the panhandle of Florida and spent most of his formative years on the space coast. He took an interest in chemistry and biology at a young age and was very curious about the nature of life around him. He had a particular fascination with the plants in his father's garden; examining young pumpkin sprouts, propagating new aloe vera cuttings, and examining their citrus trees. By age 17, James was engrossed by questions related to modern agricultural systems. He was drawn to topics such as mycology, hydroponics, and alternative cultivation systems. He studied plant science at the University of Florida where he became increasingly interested in how crops could be improved and the role of statistical methodology in plant breeding. This led to his enrollment at the University of Edinburgh where he studied quantitative and population genetics. Here he engaged with probability theory, complex genetics, and biometrical methodology which he then applied to studying population and selection dynamics of a maize breeding population for his Msc. He then worked at Scotland's Rural College as a research assistant conducting population genetic research in barley and wheat. He subsequently applied for a PhD within Fred van Eeuwijk's group at Wageningen University. It was here James developed a keen interest in potato genetics which then led to the research contained in this thesis.

In his free time, James enjoys playing games and sports with his daughter. In whatever time remains, James loves playing piano and singing. His other hobbies include cooking, brewing, gardening, and fermenting. He is a tinkerer and passionate proponent for free and open source software, especially as it concerns open and reproducible research.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 30 ECTS (= 22 weeks of activities)

Review/project proposal

- Exploring Genetic Progress in Potato | 4.5 ECTS

In-depth / Topical / On-site Postgraduate Courses

- Crop physiology and climate change | PE&RC / University of Florida | 2019 | 2 ECTS
- Genomic prediction in animal plant breeding | WIAS | 2022 | 1.5 ECTS

Methodological / Statistical Postgraduate Courses

- Bayesian Data Analysis | PE&RC | 2024 | 1.5 ECTS
- Matrix Algebra for Engineers | Coursera | 2020 | 0.5 ECTS
- Julia Programming Language: From Zero to Expert | Udemy | 2021 | 0.5 ECTS

Invited review of journal manuscripts

- Theoretical and Applied Genetics | Applied crop genetics | 2024 | 1 ECTS

Competence, skills and career-oriented activities

- Scientific Integrity | WGS | 2023 | 0.6 ECTS
- Writing Week | PE&RC | 2023 | 1 ECTS
- Time & project management | Solynta | 2022 | 0.4 ECTS

Scientific Integrity/Ethics in science activities

- Ethics and Animal Science | WGS | 2023 | 0.8 ECTS

PE&RC Retreat, PE&RC Day, and other PE&RC events

- PE&RC weekend for first year PhD candidates | 2019 | 0.9 ECTS
- PE&RC weekend for last year PhD candidates | 2023 | 0.6 ECTS

National/local scientific meetings, seminars, and discussion groups

- Book club reading Probability Theory and Statistical Inference by Aris Spanos (Discussion over 10 chapters) | 2020 | 1 ECTS
- Department book reading of Statistical Principles for the design of experiments by Mead, Gilmour, and Mead (15 discussions) | 2023 | 1.5 ECTS
- Literature Discussions over potato related topics with industry partner (Two years participation) | 2020-2022 | 3 ECTS
- 22nd European Association for Potato Research meeting | 2024 | 1.2 ECTS

International symposia, workshops and conferences

- 21st Eucarpia general congress and potato meeting | Online due to Covid | 2021 | 2.2 ECTS - 31st Plant & Animal Genome Conference | San Diego, USA | 2024 | 2.8 ECTS

Lecturing/Supervision of practicals/tutorials

- Statistics 1b Practical | 2020 | 6 days | 1.8 ECTS
- Statistics 1a Practical | 2021 | 6 days | 1.8 ECTS
- Statistics 1b Practical | 2021 | 6 days | 1.8 ECTS

BSc/MSc thesis supervision

- Research topic 1: Novel prediction methods for pre-breeding program in potato | 3 ECTS

Acknowledgements

I begin by dedicating this thesis to the memory of my father and little sister, **Shawn & Erika Adams**. The magnitude of your absence has only become stranger as time lengthens; I miss you both. Quoting David Lynch, *"you have gone to whatever the next place is ... Death is just a change"*[1].

I want to thank all of my family for the strength you have given me through the research and writing of this thesis. Specifically, I want to thank **Flora & Jane**. You have both been so patient through this process and gave me the time I needed to do it. Without your sacrifices, it would not have been possible. I also want to thank **my mom** for her incessant encouragement through the years. Your optimism knows no bounds.

I am incredibly grateful to my supervisory committee. **Fred**, thank you for giving me the opportunity to be a member of this project. Through our many conversations, you really challenged me in how I communicate and express non-trivial statistical concepts. I would not be the biostatistician that I am now without your supervision. **Chaozhi**, I really appreciated your supervision as well. Your insight into molecular marker data and pedigree information is invaluable and I want to thank you for your crash course in all things `RABBIT`. Lastly, I want to thank **Michiel** for his immense effort as a daily supervisor. You effectively midwifed this thesis and offered so much of your time providing feedback, revision help, and unsolicited personal development advice. You were the older brother I never wanted. Thank you for that.

I want to acknowledge my previous colleagues at the WUR beginning with **Kevin** and **Henry**. The Spanos book club was the most unhinged introduction to probability theory I could have asked for. The frustration of each session was paired perfectly by the socializing that came after. Thanks to **Wenhao** for being such a great colleague. I benefited greatly from our conversations, your writing, and code sharing related to the use of identity-by-descent probabilities. Thanks to my officemate, **Jip**, for being a warm presence at Biometris. And thank you to my *buddy* **Patrick** who was a wonderful friend when I was new to the department. I appreciated the time that we shared and reflect fondly on our conversations. I am also thankful for all the administrative help from **Dinie**. Whether it was helping with enrollment, conference booking, or the IT department threatening to delete my WUR account, you were always there to help; thank you! Additional thanks to **Bart-Jan**, **Saskia**, **Ron**, **Evert Jan**, **Nienke**, **Joao**, **Martin**, **Eric**, & **Sabine** for being so approachable during my time at Biometris.

I would also like to thank all those at Solynta who made this work possible. Beginning with **Pim** for the original onboarding and introduction to all things diploid potato. Many thanks to **Ate** who was a formidable sparring partner and one of my first colleagues at Solynta. You offered a window into potato genomics and a means of making sense of it. Next, I need to thank

[1] From the late Kaleb Horton: `https://kalebhorton.ghost.io/death-is-just-a-change/`

Xarain, **Jarno**, and **Edwin** for all their work designing and executing field trials. So much of my analyses would not have been possible without your efforts. I also want to thank **Jolanda** and **Menno** for your insight into the breeding process. Moreover, I am thankful to consider you both friends. An additional thank you to **Su** for being such a good collaborator. Your questions about genetics often directed me to consider breeding problems from an alternative perspective. Thanks to all the other external PhDs at Solynta, including **Olivia**, **Julia**, and **Ernst**. Whenever we wrote together, it was easier. A special thanks to **Hein** and **Annemieke** for much of the initial administrative help when I immigrated to The Netherlands. When I relocated, I didn't have housing, I was far from family, and I was quickly going blind. Without your help arranging health insurance and visas, I would have been in a precarious position. I want to also acknowledge those people that made life here so much brighter, including **Fini**, **Dorien**, **Wendeline**, **Gitta**, **Ron**, **Martijn**, **Thijs**, **Iris**, **Maarten**, **Torsten**, & **Teun**.

Lastly, I want to recognize those friends who continued to keep me going. Thank you to **Johnny** and **Dorien** for all your support. **Xarain**, you have been such a positive and thoughtful presence these past seven years; your friendship has been a wellspring of goodness to me. Thank you to **Ernst** for your friendship over the years. You have always been a positive presence and someone I confide in when things were hard. Thanks to **Adria** for your constant support over the last year even an ocean away. And a final thanks to **Alejandro** and **Sergio** for being so generous opening your lives to me in these final months of writing. I can't thank you two enough.

To anyone I may have missed, thank you for the role you have played in getting me here.

Authorship Statement

Chapter 1: General Introduction. I was the sole author for this first chapter. My co-promoter, Michiel de Vries (**MV**), helped through the editing process. My promoter, Fred van Eeuwijk (**FE**), assisted by reviewing latter drafts. Generative artificial intelligence was also used. Specifically, the `deepseek-chat & deepseek-reasoner` model provided by DeepSeek (Version published June, 2025) was utilized for initial editing and review of early drafts.

Chapter 2: Little heterosis found in Diploid Hybrid Potato. I was the primary author of this chapter with initial conceptualization done by me and **FE**. I conducted the research for this chapter with regular technical input from Chaozhi Zheng (**CZ**) and **FE**. **MV** assisted with content-related feedback. **FE**, **MV**, **CZ** all served in the capacity of reviewers in the writing for this chapter.

Chapter 3: Efficient genomic prediction of yield and dry matter in hybrid potato. I was the primary author for this research chapter with the idea for the research coming from myself and **FE**. **CZ** helped me in early design and technical discussions for the research. Research and the writing of the first draft were conducted by me. **FE**, **MV**, **CZ** all assisted in the capacity of reviewers in the writing for this chapter.

Chapter 4: Digging up QTL: Evaluating the efficiency of marker-assisted selection in hybrid potato. The initial concept and conducting of research was done by me. **FE** steered initial QTL mapping efforts and helped determine which results were most relevant. **MV** helped both in examining the first results and reviewing the initial drafts of this research.

Chapter 5: Evaluating multiallelism in genomic prediction applications. The initial conceptualization for this chapter came from myself and **FE**. **CZ** provided technical assistance and supervision with the identity by descent estimation. Statistical modelling frameworks were selected by me. Generation of results and writing of the first draft was also done by me with **FE**, **MV**, & **CZ** providing assistance through the editing process.

Chapter 6: General Discussion. This last chapter was written by me. **MV** assisted in feedback over the narrative arc with helpful technical commentary provided by **FE** in later drafts.

All research contained in thesis was made possible through the financial support provided by De Topsector Tuinbouw & Uitgangsmaterialen (Project TKI-TU-18155).

Financial support from Wageningen University and Biometris for printing this thesis is gratefully acknowledged.

Cover photography by Flora Adams
Cover design by ProefschriftMaken
Printed by ProefschriftMaken
Modelling services by Xarain van Os

