UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Preservation and Improvement of Valencian Agro-diversity University Research Institute (COMAV)

Can genomic selection improve genetic gain in fruit tree breeding? Case studies in peach and apple

Master's Thesis

Erasmus Mundus Master Programme in Plant Breeding - emPLANT +

AUTHOR: Rodriguez Herrera, Joel

Tutor: Prohens Tomás, Jaime

ACADEMIC YEAR: 2025/2026

Universitat Politècnica de València

Can genomic selection improve genetic gain in fruit tree breeding? Case studies in peach and apple

Master's Thesis

Erasmus Mundus Master in Plant Breeding - emPLANT+

Author: Joel Rodríguez Herrera

UPV tutor: Jaime Prohens Tomás

INRAE tutors: Morgane Roth, Hélène Muranty, Sarah Ben Sadoun

Intake: 2024/2026

September 2026

Abstract

Apple (*Malus domestica*) and peach (*Prunus persica*) breeding programs face significant challenges, including long juvenile periods, high evaluation costs, and complex trait architectures. Genomic selection (GS) can accelerate genetic progress by enabling earlier selection, shortening breeding intervals, and improving parental selection. The effectiveness of GS depends on integrating genomic data and balancing genotyping costs with operational savings. This thesis evaluates the genetic and economic efficiency of GS in apple and peach breeding under realistic operational constraints.

A microeconomic framework was combined with stochastic simulations to compare conventional phenotypic selection with six alternative strategies integrating GS, phenotypic selection (PS), and genomic parental mating optimization. Selection schemes and costs were parameterized using information from commercial apple and peach breeding programs. The strategies were evaluated across three heritability levels using genetic gain as the outcome measure.

Across species and heritability levels, the greatest genetic gain was obtained when genomic information was used to optimize parental crosses and complement phenotypic selection. The relative advantage of GS was greatest for low-heritability traits. In apple, GS alone at the hybrid stage did not consistently outperform PS, highlighting the importance of genomic cross design. In peach, GS combined with genomic parental mating optimization achieved genetic gain comparable to or greater than PS while shortening the breeding cycle and maintaining a budget closer to the conventional scheme than the combined PS+GS strategy.

These results show GS can improve the efficiency of genetic progress in fruit tree breeding, but its value depends on species, program design, trait heritability, training population setup, and costs. Genomic tools are therefore most effective when strategically placed early for cross design and selection of complex traits, rather than used as replacements for phenotypic evaluation.

Keywords: fruit tree breeding, *Malus domestica*, *Prunus persica*, selection scheme, genomic parental mating optimization, stochastic simulation, microeconomic framework, genetic gain

Acknowledgements

I would like to express my deepest gratitude to my family and my partner for their constant support throughout my Master's degree. Thank you for accompanying me not only during this academic journey, but also through the many personal challenges, changes, and experiences that came with it. Your encouragement, patience, understanding, and presence gave me the strength and motivation to continue moving forward. This achievement is also a reflection of the support and confidence you have always given me.

I am also sincerely grateful to all the professors who contributed to my academic formation throughout the Master's program. I would especially like to thank my supervisors at INRAE, Morgane, Hélène, and Sarah, for their guidance, availability, and valuable advice throughout this work. Their thoughtful feedback, scientific expertise, and enriching discussions were essential to the development of this thesis and to my professional and personal growth. I am very grateful for the trust they placed in me and for the opportunity to learn from them.

Finally, I would like to thank all friends and colleagues who, in one way or another, accompanied me during this academic and personal journey. Your support, encouragement, and shared experiences made this path especially meaningful.

Table of Contents

ABSTRACT **2**

ACKNOWLEDGEMENTS **3**

TABLE OF CONTENTS **4**

I. INTRODUCTION **6**

1.1 KEY ASPECTS FOR APPLE AND PEACH BREEDING 6

1.1.1 Economic aspects *6*

1.1.2 Fundamental biology *6*

1.1.3 Domestication and diversity *8*

1.2 MODERN BREEDING PROGRAMS 10

1.2.1 Current challenges *11*

1.2.2 Breeding objectives *12*

1.2.3 Marker-Assisted Selection *13*

1.2.4 Genomic Selection adoption *14*

1.3 COST-EFFECTIVENESS OF GENOMIC SELECTION 17

1.3.1 Breeder's equation *17*

1.3.2 Assessment frameworks of Genomic Selection *18*

1.3.3 Cost components in economic assessments *22*

1.3.4 Effectiveness measures *23*

1.3.5 Factors affecting cost-effectiveness *24*

II. OBJECTIVES **24**

III. MATERIALS AND METHODS **25**

3.1 ECONOMIC FRAMEWORK 26

3.2 SIMULATION FRAMEWORK 27

3.2.1 Genotypic data *27*

3.2.2 Modeled trait *28*

3.2.3 Initialization of the Parental Population *29*

3.2.4 Initialization of the Training Population *30*

3.2.5 Breeding schemes *31*

3.2.6 Genomic Prediction models and cross planning algorithms *36*

3.2.7 Workflow management *37*

IV. RESULTS 42

4.1 Genetic gain in apple 42

4.2 Genetic gain in peach 43

4.3 Effect of heritability on genetic gain 44

4.4 Effect of Optimal Contribution Selection and genetic diversity 44

4.5 The added value of genomics-assisted crosses 45

V. DISCUSSION 49

5.1 Heritability determines the comparison 50

5.2 Optimizing training population to sustain prediction accuracy 51

5.3 Comparison between crops 52

5.4 Genetic gain and genetic diversity management 54

5.5 Economic interpretation and practical adoption 55

5.6 Limitations and future simulation perspectives 57

VI. CONCLUSIONS 59

REFERENCES 60

SUPPLEMENTARY MATERIAL 76

I. Introduction

1.1 Key aspects for apple and peach breeding

1.1.1 Economic aspects

Apple (*Malus domestica*) and peach (*Prunus persica*) are among the most economically important temperate fruit crops in the Rosaceae family. Apple is the third most cultivated fruit crop globally, with a total production of 97 million tons in 2024 and an estimated market value of $111 billion USD (*FAOSTAT*, 2024). China leads in production, with significant contributions from Europe and the USA (fig. 1a). Its yield has increased by approximately 60% since 1990 (fig. 1c).

Similarly, global peach production in 2024 was estimated at 27 million tons, with China as the leading producer, followed by Europe (fig. 1b; FAOSTAT, 2024). It had an estimated gross production value of about $21 billion USD and a ~52% increase in yield over the last two decades (fig. 1d; *FAOSTAT*, 2024). Beyond its agricultural value, peach also serves as a critical genetic model for the genus *Prunus* and the broader Rosaceae family (Osorio et al., 2026).

1.1.2 Fundamental biology

Both crops share a common ancestry, as they belong to the Rosaceae family and the *Amygdaloideae* subfamily (Yamamoto & Terakami, 2016). However, their fundamental biology diverges significantly. Peach serves as a key genetic model for the genus *Prunus* because of its relatively small diploid genome of 256 Mb (*Prunus Persica* | *GDR*, n.d.) and its high level of self-compatibility (Biscarini et al., 2017; Da Silva Linge et al., 2021). Conversely, apple is an outcrossing species characterized by a gametophytic self-incompatibility system and a larger heterozygous genome of 750 Mb (*Malus x Domestica* | *GDR*, n.d.), being a diploid but exhibiting ancient polyploid characteristics (A. F. Iezzoni et al., 2020; Velasco et al., 2010).

The peach fruit is anatomically classified as a drupe, characterized by a fleshy mesocarp that surrounds a hard endocarp, commonly known as a stone or pyrene, which contains a single seed (Zheng et al., 2014). Apple fruit is classified as a pome because its edible portion develops from the hypanthium, which is composed of five carpels, and surrounds

b)

a central core derived from the ovary, which contains several seeds (Dardick & Callahan, 2014; Drazeta et al., 1999).

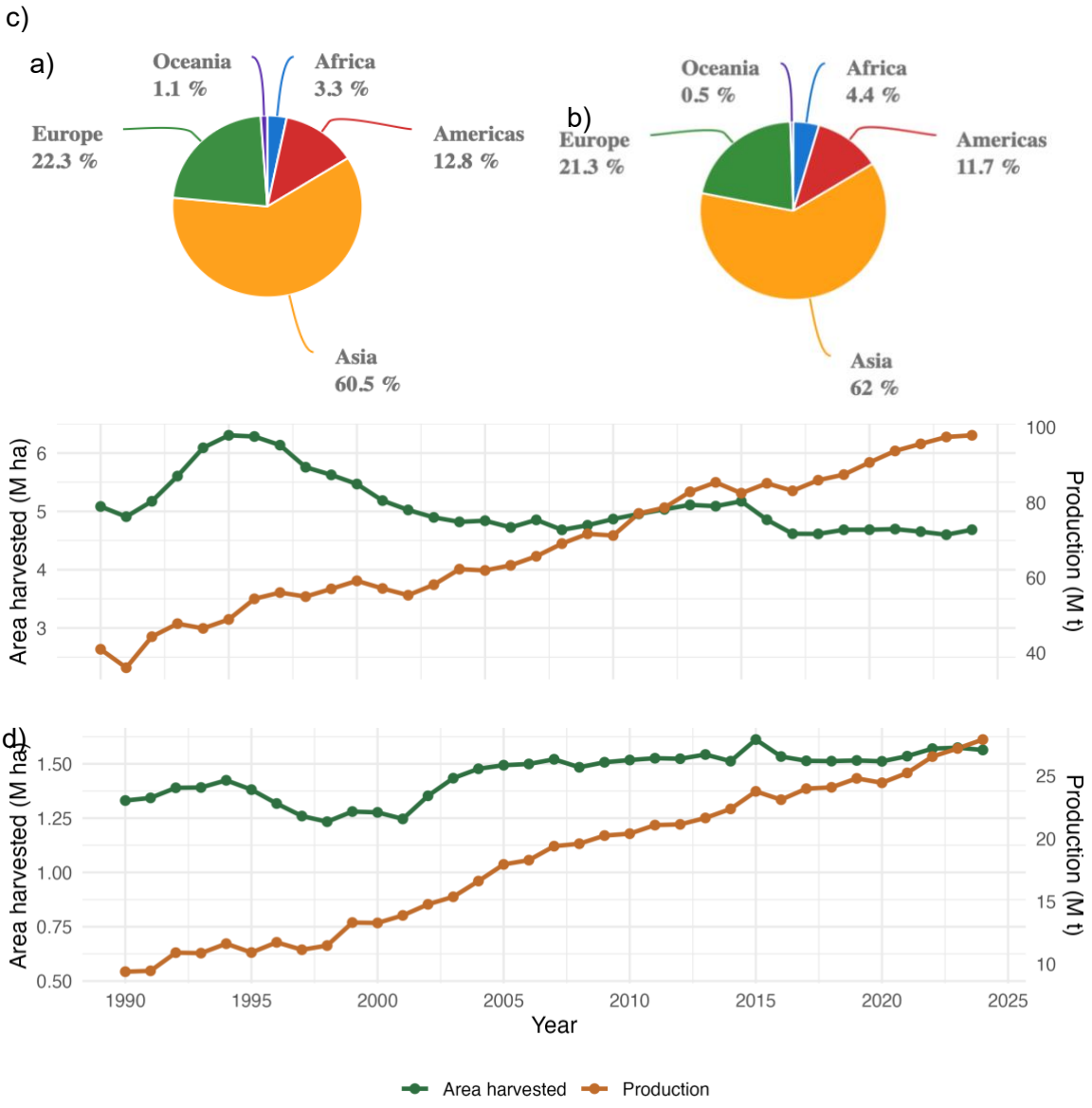


Figure 1. Total production of apples (a) and peaches (b) in 2024 by region, and their production yield since 1990 (*FAOSTAT*, 2024). China is the main contributor to the production in Asia. M ha = *million hectares*; M t = *million tons*.

The genetic improvement of these species faces challenges due to the biological constraints of perennial fruit trees. Both apple and peach trees have long juvenile periods, delaying the collection of relevant phenotypic data because breeders must wait until the trees mature to evaluate harvestable fruit traits.

Peach requires an average of three years to reach reproductive maturity (Monet et al., 1996), allowing for relatively faster breeding cycles compared to apple, which can take up to seven years to transition from seed to the reproductive phase (Kumar et al., 2012). As a result, a traditional selection cycle can last more than 20 years (OConnor et al., 2021).

Despite these long juvenile periods, both apple and peach benefit from their capacity for vegetative propagation through grafting. This propagation strategy allows breeders to extensively replicate and evaluate clones across multiple years and environments to assess genotype-by-environment (GxE) interactions and to exploit complex non-additive effects, such as dominance and epistasis, while maintaining high heterozygosity within elite genotypes (Aranzana et al., 2019; Pacheco-Ruiz et al., 2025).

The large physical size of fruit trees requires extensive experimental orchards and significant labor, which limits the number of individuals that can be evaluated in field trials. Consequently, tree spacing becomes a critical constraint in experimental designs. For example, a specific field evaluation in apple orchards employs planting spacings of 3.0 meters between rows and 0.5 meters between trees within the row (Kumar et al., 2012).

In peach, canopy dimensions vary with genetically determined tree architectures, including standard spreading forms, semi-dwarfs, and brachytic dwarfs (Aranzana et al., 2019). Although substantial genetic variability in tree architecture also exists in apple, modern commercial apple orchards are predominantly trained to two-dimensional trellised systems, whereas peach production still encompasses a wider diversity of tree architectures. Apple tree height and vigor are primarily managed through the selection of rootstocks, with dwarfing and semi-dwarfing rootstocks commonly utilized to reduce overall tree height and enable high-density plantations (Beckman & Lang, 2003).

1.1.3 Domestication and diversity

Apple was domesticated about 4,000-10,000 years ago in Central Asia (Cornille et al., 2014), and since then, thousands of apple cultivars have been propagated and conserved through grafting (Way et al., 1991). However, nowadays commercial production relies on a few high-performing clones. Over half of US apple plantings are Red Delicious, Gala, Fuji, or Honeycrisp, with ancestors like Golden Delicious, McIntosh, and Jonathan in their pedigrees (Tegtmeier et al., 2025).

Peach likely originated in southwest China, with fossils and DNA evidence indicating its emergence in the Pliocene (Y. Yu et al., 2018). Edible peaches first appeared between 3.47 and 2.6 million years ago, with fossils from Kunming, China, dating to 2.6 million years ago and resembling modern peach stones, indicating they existed before human cultivation (Y. Yu et al., 2018). Archaeological evidence from the Yangtze River valley indicates peach cultivation 8500–7500 years ago, though some suggest domestication in northwest China around 4000–5000 years ago (Y. Yu et al., 2018; Zheng et al., 2014).

Modern cultivated peach cultivars also come from a few specific founder genotypes. The “Chinese Cling” landrace is recognized as a key founder genotype in the history of global peach breeding in China, Japan, the USA, and Europe due to its highly desirable flavor and strong aroma (Li et al., 2023).

Cultivars such as the low-acid nectarine 'Big Top® Zaitabo' have become industry standards for their sweet taste and good postharvest quality (Laurens et al., 2018). The cultivar 'Lovell' is of fundamental scientific importance worldwide, as its fully homozygous doubled-haploid genotype was used to create the first high-quality reference genome sequence for the species (Verde et al., 2013).

Peach production also depends on regionally important cultivars. In Europe, and particularly in France, cultivars such as 'Coraline® Monco' and 'Big Top® Zaitabo' are widely used, while 'Elberta', 'Redhaven', and 'O’Henry' have been important in North America, 'Akame' and 'Hakuto' in Japan, and numerous landraces remain important in China (Zheng et al., 2014).

Contrasting patterns of genetic diversity are observed in apple and peach, reflecting differences in domestication history and breeding strategies. The large-scale evaluation of the collections within the FruitBreedomics project revealed that apples exhibit high levels of genetic diversity and a weak population structure, confirming prominent, large-scale gene flow across Europe (Laurens et al., 2018). Conversely, peach cultivars exhibited an unbalanced distribution of homozygous and heterozygous regions along the genome, characterized by large monomorphic regions that are identical by descent due to a remarkably high level of co-ancestry (Laurens et al., 2018).

1.2 Modern breeding programs

Standard fruit tree breeding generally follows three main stages (fig. 2). First, suitable parents are identified and crossed, and the best offspring are selected from large families. Since the parents are not pure lines, these hybrids should be called ‘pseudo-F1’, but for simplification, we later use the term F1. Because these species are highly heterozygous and clonally propagated, each F1 seedling undergoes preliminary screening in unreplicated trials. Selected individuals are then vegetatively propagated. These clones are advanced to multi-year trials across diverse environments to evaluate phenotypic performance, yield stability, and GxE interactions. Finally, the elite advanced selections undergo larger-scale testing, often conducted within commercial orchards to confirm agronomic and economic viability (Bančič et al., 2025; OConnor et al., 2021).

Thus, the first stage of each selection cycle involves a large-diversity phase, which can be the most costly, both economically and in selection effort, due to large family sizes. However, this stage also offers the best opportunity for optimization through early and more precise identification of promising genotypes (Muranty et al., 2015).

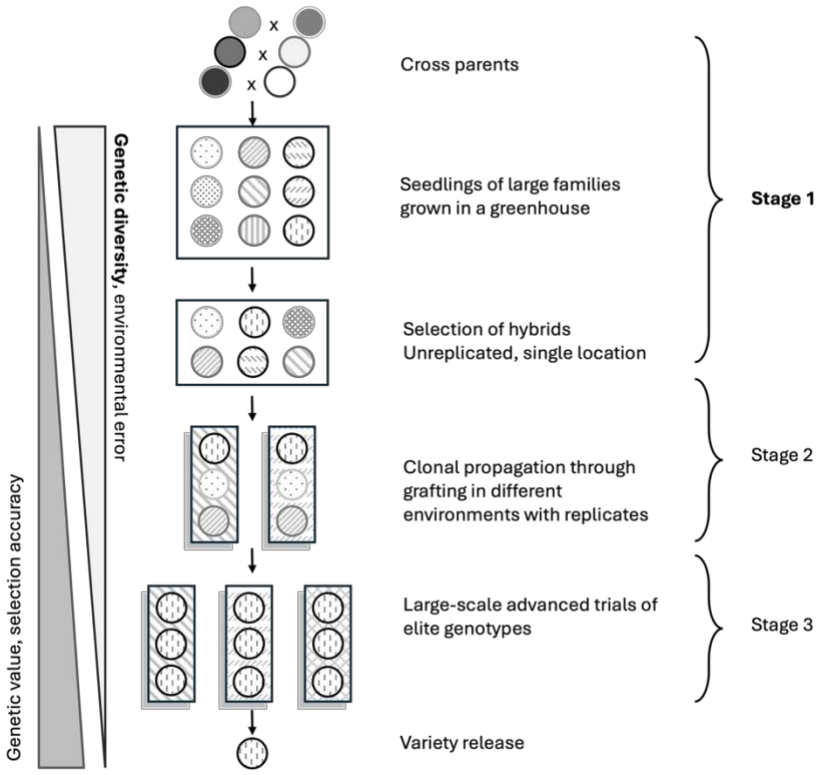


Figure 2. General selection scheme exemplifying a typical clonally-propagated tree breeding program, including a first, large-diversity phase of unreplicated plants and two stages of clonal selection. The first phase is when MAS is commonly used and GS has greater potential. Planting and growing periods should be considered between each selection campaign.

1.2.1 Current challenges

Despite differences in germplasm genetic diversity, both species face similar breeding challenges associated with the repeated use of elite founders. Extensive domestication and recurrent selection for desirable horticultural traits have reduced genetic polymorphism and increased co-ancestry within modern commercial germplasm pools (Migicovsky et al., 2021; Shimizu & Nonaka, 2025; Urrestarazu et al., 2016).

Many current commercial cultivars were also developed for production systems heavily reliant on chemical pesticides, leaving them highly susceptible to prevalent endemic diseases (Aranzana et al., 2019; Jung et al., 2022). As production systems shift toward more sustainable and pesticide-free approaches, this restricted genetic diversity limits the exploitable variation available to breeders and constrains the development of cultivars resilient to both existing pathogens and emerging biotic and abiotic stresses (Aranzana et al., 2019; Serrie et al., 2024).

Moreover, some important agronomic traits, such as fruit weight, yield, and soluble solids content, are highly polygenic quantitative traits. These traits often exhibit low-to-moderate narrow-sense heritabilities (h^2) and are strongly influenced by environmental factors and G×E interactions, which can make traditional phenotypic selection (PS) challenging. However, GxE interactions also highlight the importance of local adaptation, suggesting that breeding objectives should not necessarily focus on developing cultivars that perform uniformly across a wide range of environments, but rather on cultivars adapted to specific environmental conditions. Such a strategy may also promote greater diversity among cultivated varieties (Serrie et al., 2024).

Although these challenges can be partially offset by vegetative propagation, grafting simultaneously complicates selection by introducing rootstock-scion interactions (Mauro et al., 2022) and preventing the precise evaluation of selection candidates on their own roots (Kumar et al., 2011).

Despite biological bottlenecks, boosting breeding efforts is essential to address agronomic concerns and meet market preferences. Consumer expectations demand higher nutritional content and fruit organoleptic quality. Thereby, consumption can be reduced by inconsistent fruit quality (Cirilli et al., 2016; Pacheco-Ruiz et al., 2025; Peil et al., 2011).

Futhermore, the long juvenile phases and the need for prolonged, multi-year field evaluations to obtain reliable phenotypic data at physiological maturity (Magon et al., 2023; Pacheco-Ruiz et al., 2025; Shimizu & Nonaka, 2025) limit traditional PS, which remains slow and resource-intensive, severely limiting the rate of genetic gain and often requiring decades to deliver new commercial cultivars (Robinson et al., 2025).

1.2.2 Breeding objectives

Currently, fruit quality and yield are the main breeding goals in commercial apple and peach production, with disease resistance being important but secondary (Peil et al., 2011; Serrie et al., 2024). Peach breeding programs primarily focus on improving fruit quality to meet market demands, including traits such as longer storability, transport resistance, specific shapes, and better nutritional value (Lirong et al., 2020). Surveys show that producers prioritize fruit quality traits over disease resistance because they directly affect consumer acceptance and retail value (Yue et al., 2013). The importance of these traits varies by location and experience, influenced by economic factors such as orchard costs and consumers' willingness to pay for high-quality fruits (Yue et al., 2013).

Apple producers prioritize fruit flavor and crispness as the most important traits (Yue et al., 2013), alongside scab resistance (Laurens et al., 2018), shelf life, and firmness (Yue et al., 2013). Peach breeding centers on enhancing sugar content and sweetness perception (Cirilli et al., 2016), with additional emphasis on bacterial spot resistance (Gasic et al., 2015), post-harvest shelf life, and fruit appearance (A. Iezzoni et al., 2016).

Developing varieties with predictable blooming times suited to specific environments is key to adapting crops to different climates. Global warming often accelerates plant growth and causes plants to bloom earlier, increasing the risk of spring frost damage (Aranzana et al., 2019; Cazenave, 2022). Breeders systematically evaluate traits like bloom and ripening dates (Da Silva Linge et al., 2021). As climate change continues to impact flowering and productivity, creating new, climate-resilient cultivars remains a priority (Peil et al., 2011).

In peach breeding, defining targets requires considering fruit classes because market segments demand different traits, such as peach or nectarine type, flesh color, and fruit shape. These combinations create profiles essential for meeting consumer needs. A similar segmentation based on visual attributes also occurs in apple breeding, where producers and breeders target specific skin ground and over colors, such as distinct red or green varieties (Jung et al., 2022).

Breeders must develop cultivars covering the entire production calendar, with overlapping harvest periods to fit narrow optimal harvest windows. This is particularly critical in peach, where fruit generally has a much shorter storage life than apple, making the precise positioning of cultivars within the harvest calendar essential for producers. If harvests are excessively concentrated within the same period, temporary oversupply can also lead to price reductions.

Therefore, extending the overall ripening season and filling specific gaps in the production calendar are major strategic objectives. Breeders must carefully evaluate maturity dates for crop scheduling and target specific production time slots that complement the available cultivar catalog (Frett et al., 2012; Osorio et al., 2026).

Tree architecture, although less often discussed than fruit quality, is important for maximizing yield and reducing orchard management costs. Therefore, managing vegetative development is essential for controlling tree geometry and facilitating orchard operations, and it requires specific pruning efforts and training systems (C. Zhang et al., 2023).

Key traits include interconnected characteristics that affect agronomic performance, fruit quality, and market acceptance, grouped into productivity and organoleptic quality. These traits form an integrated framework for early elimination of undesirable seedlings and advancement of promising ones, influenced by heritability, phenotypic variation, and prediction accuracy (Muranty et al., 2015).

1.2.3 Marker-Assisted Selection

Breeding programs employ complementary conventional and molecular approaches, such as Marker-Assisted Selection (MAS; Table 1). However, the effectiveness of molecular methods varies substantially. Single-gene disease resistances show high success with MAS (Iezzoni et al., 2020), whereas quantitative traits such as sugar content remain challenging due to low heritability and environmental influence (Cirilli et al., 2016).

MAS is crucial for pyramiding disease-resistance genes into elite germplasm to confer durable resistance. Large-scale apple breeding programs have developed pre-breeding materials by combining up to seven resistance genes, targeting apple scab (Rvi2, Rvi4, Rvi6) and powdery mildew (Pl1, Pl2, Pld, Plm; Laurens et al., 2018). Specific crossing

schemes have accumulated up to three scab resistance genes (Rvi2, Rvi4, Rvi6) and generated multi-disease pyramids coupling powdery mildew (Pl1, Pl2) with scab (Rvi4, Rvi6; Peil et al., 2011). Integrating MAS with rapid breeding has accelerated the stacking of resistance alleles, at Rvi6, RviHC, FB5, and FB7, from multiple donors (A. F. Iezzoni et al., 2020).

MAS can also be combined with strategies that shorten the interval between generations. The low-input fast-track (LIFT) system, for example, integrates early molecular identification of seedlings carrying target resistance genes with accelerated greenhouse growth cycles and artificial vernalization. This approach can reduce generation time to approximately two years, thereby accelerating the introgression and stacking of resistance genes in apple (Bühlmann-Schütz et al., 2025).

1.2.4 Genomic Selection adoption

Genomic selection (GS) uses genome-wide markers to predict the genetic potential of individuals, bypassing or complementing phenotypic evaluation. It relies on a reference population that is genotyped and phenotyped to build a model that estimates marker effects. This model predicts genomic estimated breeding values (GEBVs) for genotyped candidates in a target population. Breeders use GEBVs to select top individuals, speeding breeding cycles and boosting genetic gain (Budhlakoti et al., 2022).

GS can be particularly promising for perennial Rosaceae crops, where many breeding targets are governed by complex polygenic architectures (Pacheco-Ruiz et al., 2025; X. Zhang et al., 2025). In these species, high heterozygosity generates broad phenotypic variation and allows favorable dominance effects to be maintained through clonal propagation, thereby supporting continued genetic improvement. At the same time, this genetic complexity complicates parent selection and the prediction of superior crosses, especially for low-heritability traits such as fruit size, which is also strongly influenced by environmental conditions and crop management.

To address these limitations, modern breeding programs are increasingly shifting from predominantly empirical PS toward data-driven approaches that integrate genomic information. Within this framework, GS can improve the prediction of breeding values and support more informed parental choice and cross design for complex traits (Pacheco-Ruiz et al., 2025).

Table 1. Representative applications of MAS in apple and peach breeding programs.

Species	MAS application	Representative loci/markers	Main use	References
Apple	Disease resistance	*Rvi6 (Vf)*, *FB_MR5*, *Pl1*	Early seedling selection and pyramiding of resistance against apple scab, fire blight, and powdery mildew	(Baumgartner et al., 2016) (Iezzoni et al., 2020)
Apple	Fruit texture and storability	*MdACS1*, *MdACO 1*, *MdPG1*	Prediction of firmness, ripening behavior, texture, and postharvest storage performance before fruiting	(Laurens et al., 2018); (Kumar et al., 2012)
Apple	Fruit color and acidity	*MdMYB10*, *Rf*, *Ma* QTL	Selection for red skin coloration and fruit acidity balance	(Iezzoni et al., 2016)
Peach	Disease resistance	*PpeXapF*, *Xap.Pp. OC-1.2*, *Xap.Pp.OC-6.1*	Early screening for bacterial spot resistance using SNP and KASP markers	(Iezzoni et al., 2020); (Miller et al., 2019)
Peach	Fruit commercial type	*G*, *Y*, *F-M*, *D* loci	Prediction of nectarine skin, flesh color, melting/freestone texture, and low acidity before field planting	(Laurens et al., 2018); (Iezzoni et al., 2020)
Peach	Fruit quality and maturity	SNP and SSR markers linked to maturity date, soluble solids content, and texture	Early selection for sweetness, harvest timing, and texture-related quality traits	(Da Silva Linge et al., 2021); (Pacheco-Ruiz et al., 2025)

Unlike traditional MAS, which targets a few major-effect genes, GS uses dense genome-wide molecular marker datasets to simultaneously estimate the effects of all QTLs (Magon et al., 2023; W. Yu & Cheng, 2024). These high-density datasets can be obtained through various high-throughput genotyping platforms, including not only standard SNP arrays but also next-generation sequencing approaches such as genotyping-by-sequencing (GBS) and whole-genome sequencing (WGS) (Seyum et al., 2022).

Using dense datasets demands significant computational resources and poses statistical challenges. When the number of molecular markers far exceeds phenotyped individuals, overparameterization can cause overfitting and poor predictions. Dense genome coverage may also lead to multicollinearity, as many markers are correlated due to linkage disequilibrium (LD; Budhlakoti et al., 2022; Seyum et al., 2022).

To address statistical challenges, GS uses specialized mathematical frameworks that handle high-dimensional data without relying on strict prior marker selection. Predictive

pipelines employ penalized regression models such as ridge regression and Bayesian approaches that assign prior distributions to marker variances (Budhlakoti et al., 2022).

Calculating GEBVs at the seedling stage allows breeders to make informed selection decisions before plants reach maturity (Kumar et al., 2012; Li et al., 2023). This approach fundamentally reconfigures the breeding pipeline by influencing all parameters of the breeder's equation (eq. 1), which defines genetic gain over time (Campoy et al., 2016; Kumar et al., 2012). GS can reduce the generation interval, enable higher selection intensities, improve prediction accuracy for complex traits, and allow precise monitoring of additive genetic variance and inbreeding (Muranty et al., 2015; Wartha & Lorenz, 2021).

Empirical studies and simulations in rosaceous crops have consistently demonstrated the substantial potential and high predictive abilities of GS for various complex traits. In apple breeding, prediction accuracies for fruit quality traits such as firmness and sugar content have been reported to range from 0.70 to 0.90 (Kumar et al., 2012). These results were obtained using a training population of approximately 1,200 individuals derived from several F1 breeding families, representing a relatively favorable scenario for genomic prediction.

In European peach populations, average predictive abilities of 0.65 were reported, reaching 0.84 for fruit weight and 0.83 for acidity. However, these optimistic values were strongly influenced by experimental designs in which predictions were evaluated within closely related full-sib families (Biscarini et al., 2017).

Furthermore, this early prediction capability establishes two major complementary routes for breeding efficiency: it not only facilitates the early selection of promising candidates for advanced testing as potential commercial cultivars, but it also enables the rapid reuse of these superior candidates as next-cycle parents (Kumar et al., 2012). This drastically reduces the generation interval and accelerates the accumulation of favorable alleles in the parental pool, boosting the overall rate of genetic progress across cycles (Kumar et al., 2020).

In this way, genomic-based decisions can also support parental mating designs at the crossing stage starting each cycle. Genomic Predicted Cross-Performance (GPCP) is an emerging approach that optimizes parental combinations by predicting the expected total genetic value of their progeny, rather than relying solely on the GEBVs of individual parents (Nyaga et al., 2025). GPCP is particularly beneficial in clonally propagated crops prone to

inbreeding depression because it exploits additive and dominance effects to maximize heterosis through parent complementarity (Nyaga et al., 2025; Werner et al., 2023).

Another important aspect for breeders is resolving the trade-off between short-term genetic gain and long-term genetic diversity. Optimal Contribution Selection (OCS) is integrated into GS frameworks as an optimization strategy to balance maximizing genetic gain with preserving genetic diversity (Allier et al., 2019). While GPCP maintains more heterozygosity and dominance variance, coupling it with OCS sustains the breeding pipeline by reducing inbreeding and limiting elite genetic overrepresentation (Vieira et al., 2025). This prevents rapid genetic erosion and keeps GS a durable strategy.

Nevertheless, general operational adoption of GS still depends on its economic viability and cost-effectiveness. The integration of high-throughput genotyping requires significant capital investment, introducing new financial dynamics into established breeding operations (Heffner et al., 2009; Iwata et al., 2016). The primary economic benefit of GS derives from reduced phenotyping and field maintenance costs, as genetically inferior seedlings can be discarded before being grafted and planted in orchards.

1.3 Cost-effectiveness of Genomic Selection

1.3.1 Breeder's equation

The theoretical efficiency of a breeding program is quantified by the breeder's equation (Lush, 1937; eq. 1), which defines the expected response to selection:

$$R = h^2 S$$

Equation 1, where:

R is the response to selection;
h^2 is the narrow-sense heritability;
S is the selection differential.

And derives the equation for genetic gain per time (eq. 2):

$$\Delta G = \frac{i \cdot r \cdot \sigma_A}{L}$$

Equation 2, where:

i is the selection intensity;
r is the selection accuracy;
σ_A is the genetic variance;
L is the generation interval.

These equations provide an essential framework for breeders, helping them optimize strategic decisions and identify multiple pathways to maximize genetic progress while efficiently managing operational costs and resource allocation (Vieira et al., 2025; Wartha & Lorenz, 2021).

GS can boost genetic gain by reducing the generation interval through early DNA-based evaluation, increasing selection intensity by discarding inferior seedlings before orchard planting, and improving selection accuracy using genome-wide markers, and optimizing the use of genetic variance through informed selection and mating decisions.

1.3.2 Assessment frameworks of Genomic Selection

Current frameworks for assessing GS implementation fall into three non-mutually exclusive primary categories: 1) analytical frameworks comparing GS to conventional selection approaches purely based on genetic values or other selection outcomes (Heffner et al., 2010), 2) economic efficiency comparison models evaluating breeding schemes under certain investment conditions (Fugeray-Scarbel et al., 2022), and 3) constraint optimization frameworks combining selection theory with numerical optimization (Riedelsheimer & Melchinger, 2013). These frameworks can share common effectiveness measures, including genetic gain per unit of time and/or cost, return on investment, or increase in selection accuracy, but differ in their implementation perspective.

Among the simplest comparative approaches, analytical frameworks directly evaluate the expected advantages of GS over conventional breeding methods, mainly by quantifying genetic gain using deterministic or stochastic models. Both of them are based on quantitative genetics equations, including the breeder's equation (eq. 1).

1.3.2.1 Deterministic models

Deterministic models incorporate population parameters, such as heritability, selection intensity, accuracy, genetic variance, and generation interval, directly into classical forms of the breeder's equation, allowing calculation of expected genetic gain without simulating individual samples. Heffner et al. (2010) used this general framework to compare expected gain per cycle, per year, and per unit cost under conventional and genomic breeding scenarios.

However, applying this mathematical framework requires fundamental hypotheses and assumptions about the characteristics and genetic architecture of both the training and target populations. Specifically, estimating the prediction accuracy parameter in the equation requires explicit assumptions about the training population (TP) size, phenotypic quality, and degree of genetic relatedness to the target selection candidates (Ben-Sadoun et al., 2021; Wartha & Lorenz, 2021). There are also limitations for representing finite-population sampling error, Mendelian segregation, multistage selection decisions, or inbreeding accumulation without added approximations (Hassanpour et al., 2023).

Despite the possible limitations of these assumptions, deterministic frameworks provide significant advantages in terms of simplicity, interpretability, and, mainly, computation costs. Because each numerical compilation requires minimal processing time, breeders can rapidly and efficiently compare a vast array of breeding scenarios (Fugeray-Scarbel et al., 2022).

1.3.2.2 Stochastic models

To overcome the limitations of deterministic models, simulation-based approaches have become the standard for building stochastic genetic models. Instead of inputting the population parameters directly into the breeder's equation-derived formulas, they simulate individual genotypes and phenotypes, Mendelian segregation, recombination, environmental variation, genomic prediction, selection, and mating over successive breeding cycles (R. C. Gaynor et al., 2021).

True breeding values (TBV) are commonly generated as the sum of simulated QTL effects (eq. 3), phenotypes are obtained by adding environmental deviations, and GEBVs are calculated using models such as RR-BLUP or GBLUP. Realized genetic gain is subsequently measured as the change in the mean TBV (eq. 4a) or mean genetic value

(eq. 4b) of the simulated population and is averaged across multiple simulation replicates (R. C. Gaynor et al., 2021; Liu et al., 2019; Lorenz, 2013).

$$TBV_k = \sum_{j=1}^{m} x_{kj} a_j$$

Equation 3, where x_{kj} is the genotype or allele dosage of individual k at QTL j, and $a_j$is the simulated additive effect of that QTL (Lorenz, 2013).

(a)
$$\Delta G = \frac{\overline{TBV}_t - \overline{TBV}_0}{T_t}$$

(b)
$$\Delta G = \frac{\overline{GV}_t - \overline{GV}_0}{T_t}$$

Equation 4, where TBV_t and GV_t are the mean true breeding value and genetic value at cycle t, and T_t is the total elapsed time from the baseline population to cycle t.

Simulation frameworks also allow for the splitting and management of biological effects. AlphaSim operates within the Additive, Dominance, Epistatic, and GxE (ADEG) framework, in which each trait is assigned to an additive effect and, optionally to one or more other effects. Then, the more complex trait includes all the effects and allows the calculation of the total genetic value (eq. 5).

$$GV_k(x_k, w) = \mu + \sum_{j=1}^{m} a_j x_{kj}^A + \sum_{j=1}^{m} d_j x_{kj}^D + \sum_{(j,l)\epsilon\rho} e_{jl} x_{kj}^A x_{kl}^A + w\left(\mu_G + \sum_{j=1}^{m} g_j x_{kj}^A\right)$$

Equation 5, expansion derived from the genetic value definition and effects equations (C. Gaynor, 2025), including allele dosage (Lorenz, 2013), where:

$GV_k(x_k, w)$is the genetic value of individual k in environment w;
μ is the intercept used to obtain the specified founder-population trait mean;
m is the number of QTL;
a_j is the additive effect at QTL j;
d_j is the dominance effect at QTL j;
e_{jl} is the additive-by-additive epistatic effect between paired QTL j and l;
g_j is the genotype-by-environment effect at QTL j;

μ_G is the intercept for the genotype-specific environmental slope;
w is the environmental covariate;
P is the set of QTL pairs used to model epistasis;
x_{kj}^{A} and x_{kj}^{D} are the scaled additive and dominance genotype dosages.

Gaynor et al. (2017) employed stochastic simulations of entire breeding programs over multiple cycles, using AlphaSim and R software to model the behavior of complex traits with additive, dominance, GxE, and epistasis effects (Gaynor et al., 2017). Similarly, Peixoto et al. used coalescent theory to simulate hybrid crop breeding programs under varying conditions of GxE interactions and trait heritability (Peixoto et al., 2024). These comparative simulations showed potential for its use in economic analysis, reporting ~2.47 times more genetic gain with GS than conventional programs and a return-on-investment ratio approximately 5.67 times that of the baseline scenarios (Gorjanc et al., 2017).

However, comparing schemes with varying levels of financial resources can lead to misleading conclusions about their relative efficiency in real programs. Therefore, Fugeray-Scarbel et al. (2022) established a comprehensive microeconomic framework to objectively assess the economic efficiency of alternative breeding strategies by comparing the genetic gain achieved under equivalent total investments (Fugeray-Scarbel et al., 2022).

The framework requires that any additional costs of adopting new technologies, such as high-throughput genotyping, be fully offset by reducing other operational expenses, such as the number of candidates requiring costly phenotyping. This approach models breeding as an optimization problem, in which decision variables are dynamically adjusted to maximize genetic gain while accounting for the unit costs of all essential breeding operations (Fugeray-Scarbel et al., 2022).

The first step is to describe the breeding schemes being compared by simplifying the current baseline program, identifying basic operations, detailing alternative schemes, and formulating total cost equations. The second step involves evaluating unit costs for each operation by considering expenses such as staff, consumables, rentals, services, and equipment depreciation, and then compiling the final unit costs. The third step is to define the scheme dimensions for objective economic assessment, including strategies, constraints, and progeny numbers. The final step models the schemes for assessing

efficiency, developing genetic models and simulation pipelines to evaluate genetic gain, and providing practical implementation feedback (fig. 3; Fugeray-Scarbel et al., 2022).

AlphaSimR, the R library implementation of AlphaSim, provides a flexible environment for modeling complex selection designs (Gaynor et al., 2021). This makes AlphaSimR an accessible tool for implementing stochastic simulations for the final step of the microeconomic framework proposed by Fugeray-Scarbel et al. (2022).

Despite the utility of these operational and economic assessments, this type of analysis in perennial crops is not yet documented in the literature. The closest published example applying simulation-based evaluations of different GS breeding strategies in a clonally propagated fruit crop was performed in strawberry (Werner et al., 2023).

The R2D2 consortium, supported by the INRAE SELGEN metaprogram, also contributed to developing frameworks for GS in plant and animal breeding, proposing a unified model of key components for transitioning from traditional methods (R2D2 Consortium et al., 2021). Their work showed that successful implementation depends on biological constraints, organization, economics, and resource optimization across species. The project also provided an interdisciplinary platform integrating genetics, economics, and breeding strategies, supporting optimization and simulation frameworks, and even supporting this project (R2D2 Consortium et al., 2021).

1.3.3 Cost components in economic assessments

Studies incorporated different levels of detail in their cost models, with genotyping and phenotyping costs as universal components. Phenotyping costs are highly variable according to the crop and evaluation method. Genotyping costs were specified most precisely. Gaynor et al. estimated genotyping at $20 per sample (R. C. Gaynor et al., 2017), while Wartha & Lorenz reported $14 per sample (Wartha & Lorenz, 2021). Fugeray-Scarbel et al. propose two genotyping cost scenarios: 10€ and 37€ .

The cost of high-throughput genotyping is also strongly dependent on the chosen methodology. This is important because GS accuracy is also affected by the type of marker datasets (Seyum et al., 2022). Moreover, genome characteristics such as genome size and LD decay between markers should be considered when choosing the genotyping methodology.

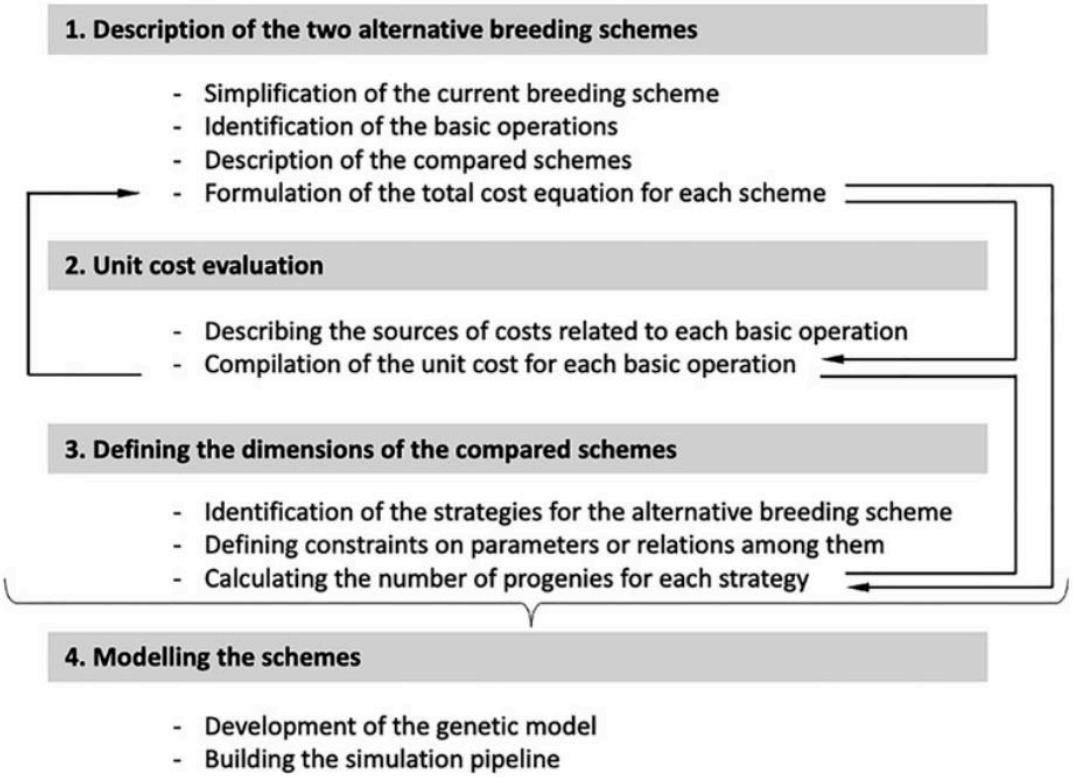


Figure 3. Workflow of the economic assessment framework proposed by Fugeray-Scarbel et al. (figure from Fugeray-Scarbel et al., 2022).

In peach, the genome size is relatively small, and LD decay is relatively slow, which allows models built on low-density SNP arrays such as the IPSC 9K SNP array V1 (Biscarini et al., 2017; Verde et al., 2012). For apple, given its larger genome size, high-density arrays could be required, and if WGS is needed to achieve adequate genome coverage, this can significantly increase GS costs.

On the other hand, infrastructure and labor costs have received limited quantitative treatment. Calvert et al. included labor and operational costs in their calculations for maximum phenotyping and genotyping capacity (Calvert et al., 2020), while Wartha et al. (2021) noted the need for advanced scientific personnel and data management infrastructure (Wartha & Lorenz, 2021).

1.3.4 Effectiveness measures

The frameworks employed multiple metrics to quantify the value of GS implementation, with genetic gain per unit time and cost as the most common measure (see Eq. 3). Gaynor et al. (2017) evaluated the two-part breeding strategy by comparing it

with conventional and standard GS strategies (R. C. Gaynor et al., 2017). Similarly, Fugeray-Scarbel et al. (2022) assessed breeding strategies constrained by fixed economic budgets, and their comparisons were purely based on genetic gain (Fugeray-Scarbel et al., 2022).

Breeding cycle time reductions provided an additional dimension for quantifying benefits. Heffner et al. (2010) concluded that GS accelerates genetic gain through its shorter breeding cycle (Heffner et al., 2010). Wartha & Lorenz (2021) even noted that reduced selection accuracy can be compensated for by a reduction in cycle time (Wartha & Lorenz, 2021).

1.3.5 Factors affecting cost-effectiveness

Cost-effectiveness depends critically on the crop, breeding scheme, secondary associated costs, and trait architecture, thereby allowing the definition of specific conditions for maximizing genetic gain. The frameworks revealed that optimal implementation strategies are budget-dependent, with constrained programs requiring low-cost imputation approaches (Gorjanc et al., 2017), moderate budgets demanding careful trade-offs between TP size and prediction accuracy (Riedelsheimer & Melchinger, 2013), and programs with more resources benefiting from combined GS and high-throughput phenotyping (Peixoto et al., 2024).

II. Objectives

The primary goal of this thesis is to provide an objective assessment of the added value of GS in peach and apple breeding. Specifically, its efficiency is compared with conventional PS in terms of genetic gain under fixed budget and operational constraints.

Moreover, this project aims to evaluate the optimal strategic placement of GS along the breeding pipeline. This involves systematically comparing GS and PS approaches at critical stages, including the choice of parental matings and the selection of hybrids during the greenhouse stage and the first observation stage.

In addition, this thesis aims to explore broader biological aspects that could influence GS suitability by comparing apple and peach breeding programs. Comparing GS

implementations across species will help identify how species-specific biological constraints influence the efficiency and economic advantages of GS strategies.

Ultimately, the analysis aims to provide breeders with objective insights into the economic trade-offs and operational constraints of integrating genomic prediction tools into the fruit breeding sector.

III. Materials and methods

The approach couples a microeconomic framework and stochastic simulations to derive comparative metrics across selection schemes. Beginning with operational parameters and operational unit costs to design budget-equivalent selection schemes, the simulations use genetic data and trait and population parameters to model several cycles of selection and estimate genetic gain.

The operational parameters are: number of stages, stage duration, number of crosses, growing periods, number of replicates per genotype, number of selected candidates per stage, stage of TP update, number and selection of new genotypes added to the TP, and parental pool renewal stage. The operational costs considered here are: crossing, greenhouse growing, plantation and field maintenance, phenotyping and genotyping. We did not account for the costs relating to scientific staff for running GS models and decision-making.

Breeders from NOVADI and CEP Innovation provided the baseline selection schemes, operational parameters, and costs for apple and peach, respectively. More details and modifications in these parameters and the selection scheme, to design alternative budget-equivalent programs, were proposed and discussed together with the breeders and collaborators from INRAE IRHS and GAFL.

The founder genotypes were built from SNP marker datasets and genetic maps. The simulated chromosomes and segregant sites were then defined from real haplotypes and genetic distances. The trait architecture defined here models only additive effects. The trait genetic mean and variance were arbitrary values for comparative reference only and do not represent a real trait.

The population parameters correspond to the sizes of the progeny population, parental pool, and TP. Because the total number of seedlings that can be produced and evaluated

is constrained by the company's operational capacity, family size was not set independently. Instead, it was calculated based on the number of crosses and the number of plants evaluated per cross. Parental pool sizes were determined by the number of founder genotypes available in each dataset. The TP size was also determined by the parental pool size and the number of new genotypes added each cycle to update the TP. Specific costs related to the parental pool and TP (*e.g.* maintenance) were not considered in the economic analysis.

An experimental grid was designed to compare the genetic gain of the baseline PS scheme with six budget-equivalent alternative scenarios that combine GS, PS, and GPCP (Table 2; page 37). Thus, GS was evaluated as an intervention that can be applied to parental and/or early hybrid selection.

For hybrid selection, a selection index (PS+GS), which weights phenotypic values and GEBVs equally, was proposed as an alternative to GS-only selection. This hybrid strategy proposes a smaller phenotyping population and fewer years of observation (Table 2) to avoid lengthening the whole cycle. In this way, the additional genotyping costs were compensated with changes in candidate numbers, phenotyping effort, stage duration, and selection intensity.

Given the biological characteristics of the crops analyzed, the simulated schemes explicitly represented the biological and operational structure of fruit-tree breeding programs, including a feasible number of crosses, family sizes, and juvenile periods. Moreover, the costs provided by the breeders also reflect different program economic dimensions.

3.1 Economic framework

The microeconomic framework proposed by Fugeray-Scarbel et al. (2022) for comparing genomic and conventional breeding schemes under equal total investment was the base analytic model (fig. 3). In this framework, the breeding program is treated as a process in which genetic gain is the output and breeding operations are the inputs. The relevant comparison is therefore the genetic gain achieved within a given budget, not the gain from adding genomic selection on top of an existing program.

For each scheme, the total cost was calculated as the sum of the number of operations multiplied by their unit costs:

$$CT = \sum_i x_i c_i$$

where x_i is the number of times operation i is performed, and c_i is the unit cost of operation i.

Applying fixed financial parameters per operation allowed the dynamic adjustment of population sizes and selection intensities across scenarios. This ensured that phenotypic and GS schemes could be compared under equivalent investment. The analysis, therefore, distinguished true improvements in breeding efficiency from apparent gains caused by increased total spending.

3.2 Simulation framework

The stochastic simulation was conducted using the R package AlphaSimR v2.1.0 (R. C. Gaynor et al., 2021). Within AlphaSimR, founder genomes were built from haplotypes and genetic maps. This ensured that the initial LD and allele frequency reflected real germplasm rather than entirely synthetic genomes.

3.2.1 Genotypic data

Apple marker data were obtained from the FruitBreedomics REFPOP apple collection high-density SNP genotyping dataset (Denancé et al., 2021; Jung et al., 2020). This dataset comprises 1,447 accessions from European institutions, genotyped with a high-density Axiom® Apple480K array (Bianco et al., 2016). 133 founder cultivars were selected based on the pedigree inferred from the same dataset (Muranty et al., 2020), retaining diploid founder (no identified parents) and semi-founder accessions (only one identified parent).

Binary SNP genotype data were parsed with PLINK v1.9.0-b.8 to recode an allele-count table. From the 320,761 SNP loci in the dataset, 24,136 markers were selected based on quality data (Howard et al., 2021; Muranty et al., 2020) and the sum of the allele dosage across accessions following the condition $0 < Allele\ count < 2 * Number\ of\ accesions$, to filter out monomorphic markers, thereby excluding loci fixed for either homozygous

genotype. Fully heterozygous loci were not excluded by this criterion, as both alleles are present and such loci can still undergo Mendelian segregation in subsequent generations.

A monotonic interpolation was used to infer the genetic positions (cM) of the selected SNP loci with only known physical positions (bp). This mapping step is required to assign markers as segregant sites for the simulation. Within each chromosome, markers with physical and genetic positions served as anchors and were first ordered by position. Genetic positions for the remaining loci were obtained via monotone cubic interpolation (Hyman method) between adjacent anchor markers, and positions outside the anchor range were clipped to the nearest boundary.

For apple, 5,146 markers with positions in the GDDH13 v.1.1 reference genome (Daccord et al., 2017) and genetic positions from the integrated genetic linkage (iGL) map, matched and updated by Howard et al. (2021), were used as anchors for the interpolation. The genetic positions of the remaining 18,990 loci were interpolated and assigned as segregant sites in the genetic map of the founder genomes (fig. Suppl. 1a).

The peach dataset consisted of a core collection of 192 accessions genotyped with the IRSC 16k Illumina SNP array (*IRSC 16K SNP Array for Prunus Persica* | *GDR*, n.d.). Wild and exotic accessions were removed, resulting in a filtered set of 168 peach varieties. The dataset contains 13,975 markers curated and imputed as described in Roth et al. (2022). After removing monomorphic loci, 13,683 markers were used as segregating sites. Marker genetic positions were inferred using 2,930 intersected loci between the consensus genetic linkage map assembled by Linge et al. (2018) and the core collection dataset with physical and genetic positions as anchors for the monotonic interpolation (fig. Suppl. 1b).

3.2.2 Modeled trait

A quantitative trait was simulated using 200 QTLs distributed proportionally across the genome according to the number of segregating sites per chromosome. SNP chips with 20K and 10K markers were simulated for apple and peach, respectively, and the marker positions were distributed proportionally in the same manner. Additive genetic variance and phenotypic mean values were set to 1 and 10, respectively. These values are only for reference and do not represent a real trait.

Since only the additive effect was modeled, the concept of heritability discussed here can be understood in the narrow or broad sense equally. Several levels of heritability were

compared, including low ($h^2 = 0.1$), intermediate ($h^2 = 0.3$), and high ($h^2 = 0.5$), to evaluate how GS outcome depends on trait architecture and trait evaluation accuracy.

For each scenario, the residual variance was fixed before the simulations using the specified additive variance (V_A=1) and target individual-plant narrow-sense heritability, setting the error variance: $V_E = V_A(1 - h^2)/h^2$. The same residual variance was used across breeding stages. When a genotype was evaluated with more clones, replication increased the precision of its mean phenotype by reducing its residual variance. Thus, the heritability of entry means increased with replication. AlphaSimR subsequently calculated realized variance components from the simulated individuals at each stage. These realized values can differ from the target because selection, drift, and the small number of individuals in later stages change the observed additive variance.

3.2.3 Initialization of the Parental Population

An initialization phase, known as burn-in, was run prior to scenario comparisons. The burn-in is recommended to build a parental population with more realistic LD patterns, population structure, and inbreeding, and to stabilize allele frequencies and genetic variances. This avoids biased results influenced by founder effects and the initial artificial random behavior. For this, the base population was initially advanced through 10 preliminary selection cycles, ensuring that the parent pool and TP represented more accurately an ongoing breeding program.

The burn-in was implemented for both species as a simple PS scheme with three cohorts (seedlings, juveniles, and mature) to represent the three years of the juvenile period (fig 4). The base population comprises 133 and 168 accessions for apple and peach, respectively. These base populations were used as first parental pools. In each cycle, 133 and 168 random crosses were made for apple and peach, respectively, to maximize the number of parents involved in crosses without forcing a specific mating plan. Family size was set to 200 for both species. The resulting seedling cohorts contained 26,600 and 33,600 genotypes for apple and peach, respectively.

The generated offspring reach the mature cohort after 3 years and are then phenotyped to select the 27 and 34 best candidates for apple and peach, respectively, corresponding to 20% of the original parental pool size and ~1/1000 of the offspring. An intrafamily selection cap was set to 2 to avoid overrepresentation of families. Finally, 80% of the best

parents are kept in the parental pool, and the selected candidates are added. In this way, the parental pool size is preserved across the burn-in cycles.

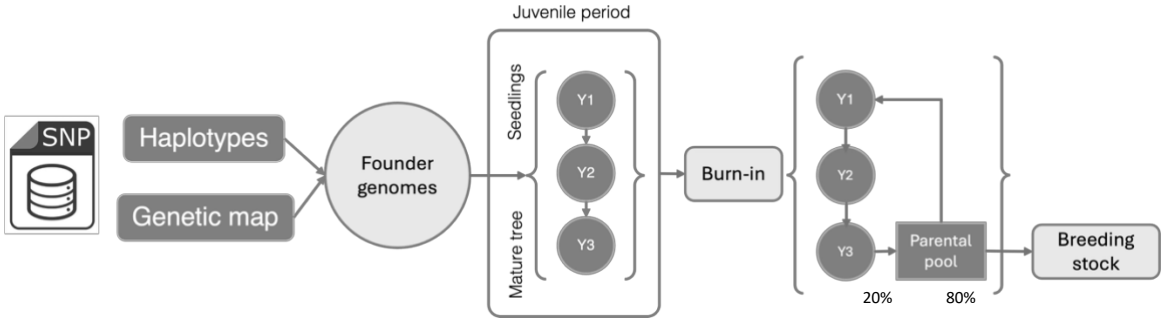


Figure 4. Burn-In scheme to initialize the parental pool, used as breeding stock, after base population creation with founder genomes. This simple selection scheme creates historical demography that represents a history of breeding rather than an artificial random initialization and leads initial transient effects to disappear from the parental pool.

The number of burn-in cycles and selection intensity were set to the minimum needed to stabilize genetic variance and inbreeding without eroding genetic diversity or exhausting future genetic gain. The last parental pool was used as material for crosses in the compared selection scheme simulations.

The same burn-in scheme and parameters were used for both species to reach fair comparative starting points and avoid influence of this phase in the comparison of final genetic gain results. This also implied using the same error variance for both species and heritability scenarios during the burn-in. Similarly, heritability in this phase was set to $h^2 = 0.1$ to avoid divergent starting points and was subsequently adjusted for the compared scenarios using the corresponding error variance.

3.2.4 Initialization of the Training Population

The parental pools from the final three burn-in cycles were combined to form the TP for genomic prediction. This yielded 399 apple and 504 peach genotype-phenotype records, corresponding to three parental pool cohorts of 133 and 168 individuals, respectively. Because parental pools were renewed across burn-in cycles by retaining 80% of existing parents and introducing selected offspring to maintain pool size, some genotypes could contribute repeated phenotype records across cohorts. For genotypes retained across multiple cycles, the same phenotype value was used in each cohort rather than generating a new observation. Phenotypes used for TP construction were those

generated during burn-in under the burn-in residual-variance setting with target h^2=0.1. The costs of constructing, phenotyping, and genotyping this initial TP were not included in the economic comparison.

3.2.5 Breeding schemes

One baseline PS scheme and six alternative genomic schemes were simulated for each species. The baseline schemes followed conventional breeding programs of NOVADI (fig. 5a) and CEP Innovation (fig. 5b) for apple and peach, respectively. The alternatives changed mating strategies, hybrid selection strategy, TP updating timing, or combinations of these components. Population sizes, stage durations, and costs were adjusted to compare schemes under similar budgets.

For both species, the breeding schemes followed the same general sequence: parental crossing, seedling growth in the greenhouse and MAS, evaluation of unreplicated F1 hybrids, replicated pre-selection trials, advanced clonal trials, and final variety release. Peach included two advanced clonal-selection levels, whereas apple included one. Each cross generated one full-sib family. Hybrid selection was subject to a stage-specific maximum number of selected individuals per family to limit the overrepresentation of individual crosses.

The simulated MAS step did not target loci affecting the modeled quantitative trait. It was therefore represented as random selection of the seedling population before the first field stage. Consequently, the comparison among scenarios concerned selection for the simulated polygenic trait and did not evaluate the effectiveness of MAS itself.

To better represent the entire breeding process, breeding cycles were modeled as overlapping pipelines. A new crossing cycle began as soon as the parental pool was renewed, while candidates from the preceding cycle continued through advanced clonal testing. Thus, the interval between crossing cycles was determined by the stage at which new parents were identified and was shorter than the time required for final variety release.

3.2.5.1 Baseline phenotypic selection schemes

The apple baseline scheme began with 15 random crosses generated from the current breeding pool. The mating maximized the use of distinct parents and prohibited

self-crossing. Genotypes were reused in different crosses only when the breeding pool did not contain the 30 distinct parental genotypes required to assign two unique parents to each of the 15 crosses. This means that only in the first cycle, starting with the 133 parental pool, were the 15 crosses made with 30 distinct parents. In subsequent cycles, the parental pool is reduced to fewer than 30 genotypes obtained via selection, and parents have to be reused for obtaining the 15 crosses.

The 15 crosses produced a target population of 10,000 seedlings, corresponding to approximately 667 offspring per full-sib family. After two years of greenhouse production, the simulated MAS step reduced this population to 3,000 hybrids. These hybrids were grafted and, after a 3-year growth period, entered the first field evaluation stage, with one clonal replicate per genotype. Each selection stage required 3 years of observation.

During the first observation stage, the number of genotypes evaluated decreased across years because of both mortality and phenotypic selection. The breeding scheme planned initially for planting 3,000 hybrids, of which approximately 2,850 were phenotyped in the first observation year after accounting for mortality. PS then reduced the population to 1,950 candidates in the second year and 600 candidates in the third year, thereby progressively reducing the phenotyping load and associated costs. For simplicity, this sequential process was not explicitly reproduced in the simulation. Instead, a single phenotypic value was simulated for each genotype, and phenotyping and selection were applied once at the end of the hybrid stage rather than over three successive observation years. Under PS, 20 genotypes were retained as pre-selections at the end of this stage.

The 20 selected hybrids were immediately used as a parental pool for the next simulated crossing cycle. Consequently, the next apple PS cycle began after completion of the first hybrid evaluation stage, nine years after the beginning of the preceding cycle, while the same candidates continued through the later stages of clonal testing. As mentioned above, because the new parental pool contained 20 genotypes, some parents could contribute to more than one cross in subsequent cycles.

The 20 pre-selections were clonally propagated and evaluated with two replicates per genotype during a six-year pre-selection stage, resulting in an evaluation field size of 40 trees while retaining 20 unique genotypes. Phenotypic selection at this stage was again subject to a family cap of one per family, retaining two advanced selections. These two candidates were subsequently evaluated with ten replicates per genotype during a further six-year advanced-selection stage. The best individual was retained as the final variety.

The complete apple baseline pipeline thus required approximately 22 years from crossing to final variety release.

The peach baseline scheme began with 20 random crosses generated using the same mating rules: self-crosses were prohibited and distinct parental genotypes were used whenever possible.

The crosses produced 4,000 seedlings, corresponding to 200 offspring per full-sib family. Following one year of greenhouse production, the simulated MAS step (random selection) retained 1,500 hybrids. These hybrids were planted with one replicate per genotype and entered the first phenotypic evaluation stage after two years of field growth. In this case, the budget considers the same population size for the three years of observation. This is one main difference from the baseline apple scheme. At the end of this stage, phenotypic selection retained 50 pre-selections.

The 50 pre-selections were clonally propagated and evaluated with two replicates per genotype during the pre-selection stage. Phenotypic selection retained eight level-1 selections. The peach parental pool was renewed at this point, one selection stage later than in apple. It consisted of the eight selected candidates and 32 additional non-selected pre-selection candidates, giving a total of 40 parents and allowing 20 crosses to be generated without necessarily reusing parents. These additional candidates were intended to be the highest-ranking remaining individuals from the pre-selection stage.

The next peach crossing cycle therefore began 12 years after the start of the preceding cycle, while the advanced trials continued. The eight level-1 selections were evaluated with three replicates per genotype per site during a five-year advanced trial, from which two level-2 selections were retained. These two genotypes were evaluated for another five years, also with three replicates per genotype per site, and the best candidate was released as the final variety. The complete peach baseline pipeline required approximately 23 years from crossing to variety release.

The main structural difference between the baseline programs was therefore not only the number of candidates and replicates, but also the timing of parental renewal. The apple program reused the 20 candidates selected after the first hybrid stage as parents, whereas peach renewed its parental pool after the replicated pre-selection stage using 40 genotypes. The peach program also included an additional advanced-selection level before variety release (see fig. 5).

3.2.5.2 Alternative Genomic Selection schemes

Six alternatives to the baseline PS scheme were evaluated, ordered by complexity:

1. GS+TP_P
2. PS+GS+TP_H
3. GPCP+GS+TP_P
4. GPCP+OCS+GS+TP_P
5. GPCP+PS+GS+TP_H
6. GPCP+OCS+PS+GS+TP_H

Here, TP_P denotes a TP updated with individuals reaching the pre-selection stage, whereas TP_H denotes a TP updated earlier with the larger set of hybrids evaluated at the first observation stage. GPCP defined the parental mating plan before crossing, and OCS restricted parental contributions to control co-ancestry. Beyond these modifications, the subsequent advanced testing stages followed the corresponding crop-specific pipelines.

In the GS+TP_P scheme, random mating was retained, but PS at the first hybrid selection decision was replaced by selection on GEBVs. Phenotypes simulated at this stage were not used to rank candidates, although they were retained to evaluate prediction accuracy.

For apple, the GS-only scheme increased the crossing stage from 15 to 17 crosses and the seedling population from 10,000 to 11,333 seedlings, thereby compensating partly for the cost of genotyping through changes in population size and later field expenses (Table 2). After MAS, 3,000 hybrids were genotyped and ranked by GEBV. A family-constrained GS retained 20 pre-selections, which immediately replaced the parental pool for the next crossing cycle, although a growth period is still required to perform the next crossing phase. These candidates were then phenotyped in the pre-selection trial, and all 20 genotyped and phenotyped pre-selections were appended to the TP for the following cycle. The next apple GS cycle could begin approximately four years after the previous crossing, and final variety release occurred after approximately 17 years (fig. 5a).

For peach, the GS scheme retained 20 crosses, 4,000 seedlings, and 1,500 genotyped hybrids. GEBV-based selection at the hybrid stage retained 50 pre-selections. As in the peach baseline, these 50 candidates were phenotyped during the replicated pre-selection stage before the parental pool was renewed. The next parental pool consisted of eight selected level-1 candidates and 32 additional candidates from the same pre-selection

cohort. All 50 pre-selections were appended to the TP because they had both marker and phenotypic records. The early parent renewal reduced the peach cycle interval to approximately eight years, and final variety release occurred after approximately 19 years (fig. 5b).

In the PS+GS+TP_H scheme, genomic information complemented rather than replaced early phenotypic evaluation. Candidates were ranked using an index based on the sum of their descending phenotypic and GEBV ranks.

- For apple, the economic design retained 15 crosses and 10,000 seedlings. A total of 3,000 hybrids were budgeted for genotyping, after which 1,050 candidates entered the reduced field evaluation cohort. These hybrids were evaluated for two years rather than the three years used in the baseline PS scheme. The 20 highest-ranking candidates according to the combined PS+GS index became both the pre-selections and the parental pool for the next cycle. All hybrids in the combined genotyped and phenotyped cohort were appended to the TP, corresponding to a TP increment of 1,050 records per cycle. The advanced stages remained unchanged, and final variety release still occurred after approximately 22 years.
- For peach, 1,500 hybrids were genotyped, of which 1,000 entered the reduced field evaluation cohort. The annual phenotyping loads subsequently decreased to 900 and 500 candidates in the next two years. This decrease was proposed to balance the budget compensating genotyping costs. The combined index was then used to identify the 50 pre-selections. As in the other peach schemes, the parental pool was renewed after the replicated pre-selection stage using eight selected candidates and 32 additional candidates. The hybrid training population (TP_H) was updated at the end of each cycle with the 500 genotyped candidates that completed the final reduced field evaluation. Thus, 500 new genotype-phenotype records were added per cycle.

In the GPCP schemes, the algorithm uses the marker effects estimated from the TP to calculate the expected performance of the progeny of the mating combinations. The downstream hybrid selection and TP-update rules were otherwise identical to either GS+TP_P or PS+GS+TP_H. The corresponding OCS variants retained the same candidate numbers, stage durations, parental-renewal points, and TP-update rules, but constrained parental contributions during cross planning.

For peach, the GPCP used the TP available in the same cycle for both cross prediction and early hybrid GS. The subsequent parental pool contained eight selected pre-selections plus 32 candidates, and the TP was enlarged by all 50 pre-selections. For apple, GPCP selected the mating plan before producing the enlarged population of 11,333 seedlings, after which 3,000 hybrids were genotyped and 20 were retained by GEBV. These 20 candidates formed the next parental pool and were added to the TP after replicated phenotyping. GPCP and OCS scenarios inherited the operational costs of the corresponding GS+TP_P or PS+GS+TP_H pipeline because the computational and personnel costs associated with cross optimization were not considered.

In the apple scenarios, different training populations were used for GPCP under the TP_P and TP_H because phenotypic information became available at different stages of the breeding cycle. Under TP_P, updating the TP for GPCP required phenotypic information from the smaller set of pre-selected genotypes, which became available relatively late in the cycle, after crosses for the next generation had already been planned. Therefore, GPCP relied on the TP available from the previous cycle. In contrast, GPCP under the TP_H could use a TP updated earlier with phenotypic information from the larger hybrid population, allowing more recent information to be incorporated at the time of selection (TP_H; Fig. 5a).

3.2.6 Genomic Prediction models and cross planning algorithms

A ridge regression best linear unbiased prediction (rrBLUP) model was used to calculate genomic estimated breeding values (GEBVs) at the first selection stages. The C++ code developed by Werner et al. (2023, https://github.com/crWerner/ClonalBreedingSim) was used to implement GPCP and OCS at the crossing stages. For OCS, a target angle of 45° was used to balance the competing objectives of maximizing short-term genetic gain and minimizing parental coancestry. In this optimization framework, a target angle of 0° places full emphasis on genetic gain, whereas 90° places full emphasis on minimizing coancestry; intermediate values define different trade-offs between these objectives. Thus, 45° was selected as a conservative intermediate setting intended to balance genetic improvement with the maintenance of genetic diversity (Cowling et al., 2023; Kinghorn, 2011; Villiers et al., 2024).

3.2.7 Workflow management

An automated workflow was implemented through Snakemake v9.16.3 (Mölder et al., 2021). Input values, such as operational and trait parameters, are defined in a YAML configuration file. Selection schemes are defined through a JSON file where stages are declared as blocks with key settings, stage operations (such as crossing or selection), and dependencies (i.e., order of operations and stages). Snakemake compiles the input values and selection schemes and builds a Directed Acyclic Graph (DAG) which guides the code execution.

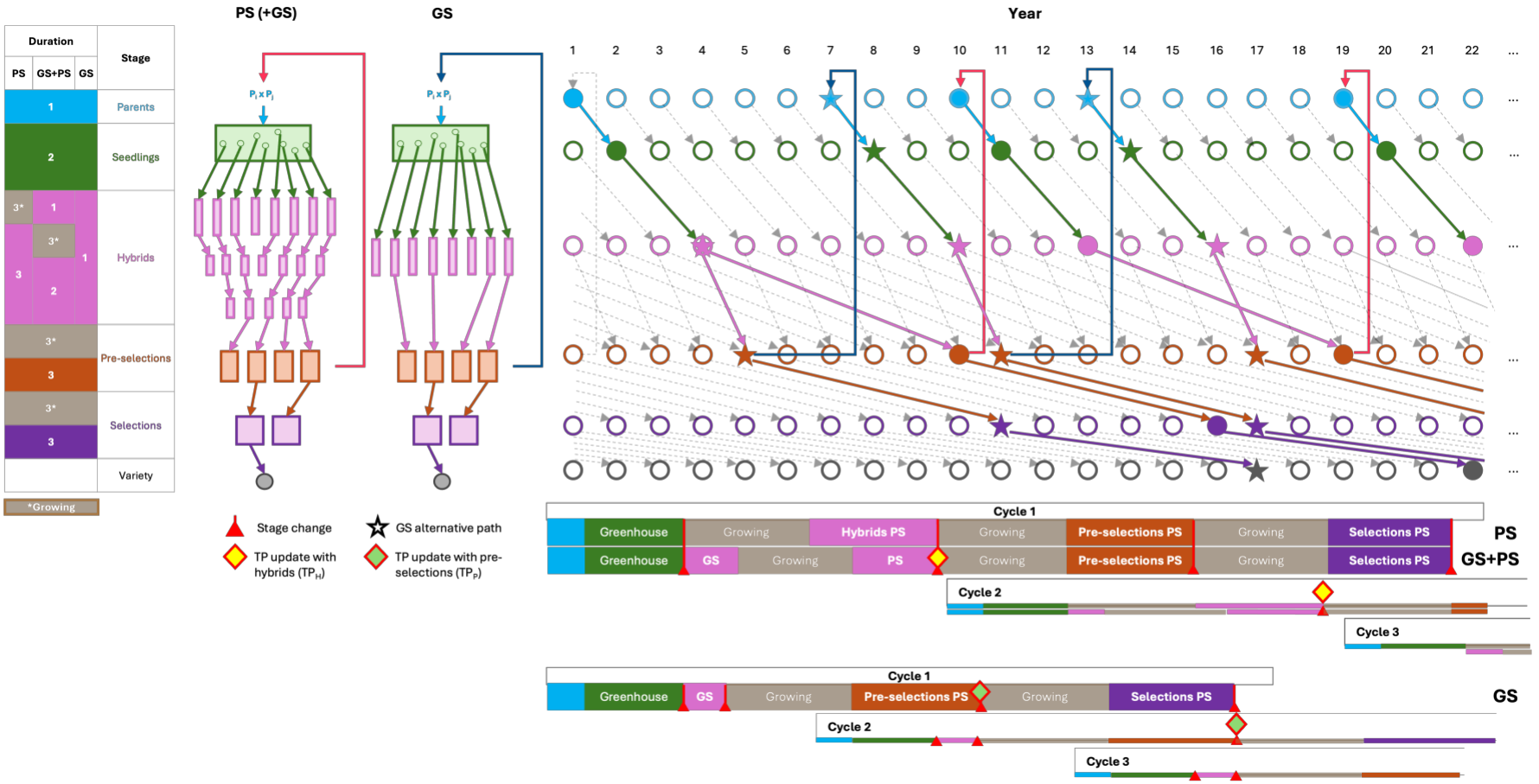
Moreover, the Snakemake workflow expands parameters and selection schemes into experimental grids, and assigns a unique scenario ID for each combination of parameters and selection schemes. It also manages replicates and parallel processes. This allows for reproducible and efficient tracking of code execution, parameters, and input and output files.

The workflow was designed to run under the SLURM workload manager through the HPC service of the Bioinfo Genotoul cluster (GenoToul Bioinfo, 2018). The execution environment was pinned through an Apptainer container to ensure stable software versions and portability.

A Shiny v1.13.0 user interface (fig. suppl. 4) developed for this project was used for workflow configuration and execution. It allows defining founder genome inputs, trait architecture, breeding scheme parameters, operational costs, and scenario variants. The Shiny server generates session-specific working directories, writes structured configuration files (YAML and JSON files), previews workflow execution with the DAG, and submits Snakemake jobs through SLURM.

The standalone version of the code is also available in the repository. These scripts replicate the same workflow but hardcode parameters, scenarios, and execution paths for each experiment. This makes them less flexible than the Shiny App but more straightforward for producing final results, as each analysis can be run directly without interactive inputs. The code repository is available upon request.

a)


Duration
PS
GS+PS
GS
Stage
Parents
Seedlings
Hybrids
Pre-selections
Selections
Variety
*Growing
PS (+GS)
GS
Year
Stage change
TP update with hybrids (TPH)
GS alternative path
TP update with pre-selections (TPP)
Cycle 1
Greenhouse
Growing
Hybrids PS
Pre-selections PS
Selections PS
PS
GS
GS+PS
Cycle 2
Cycle 3

b)

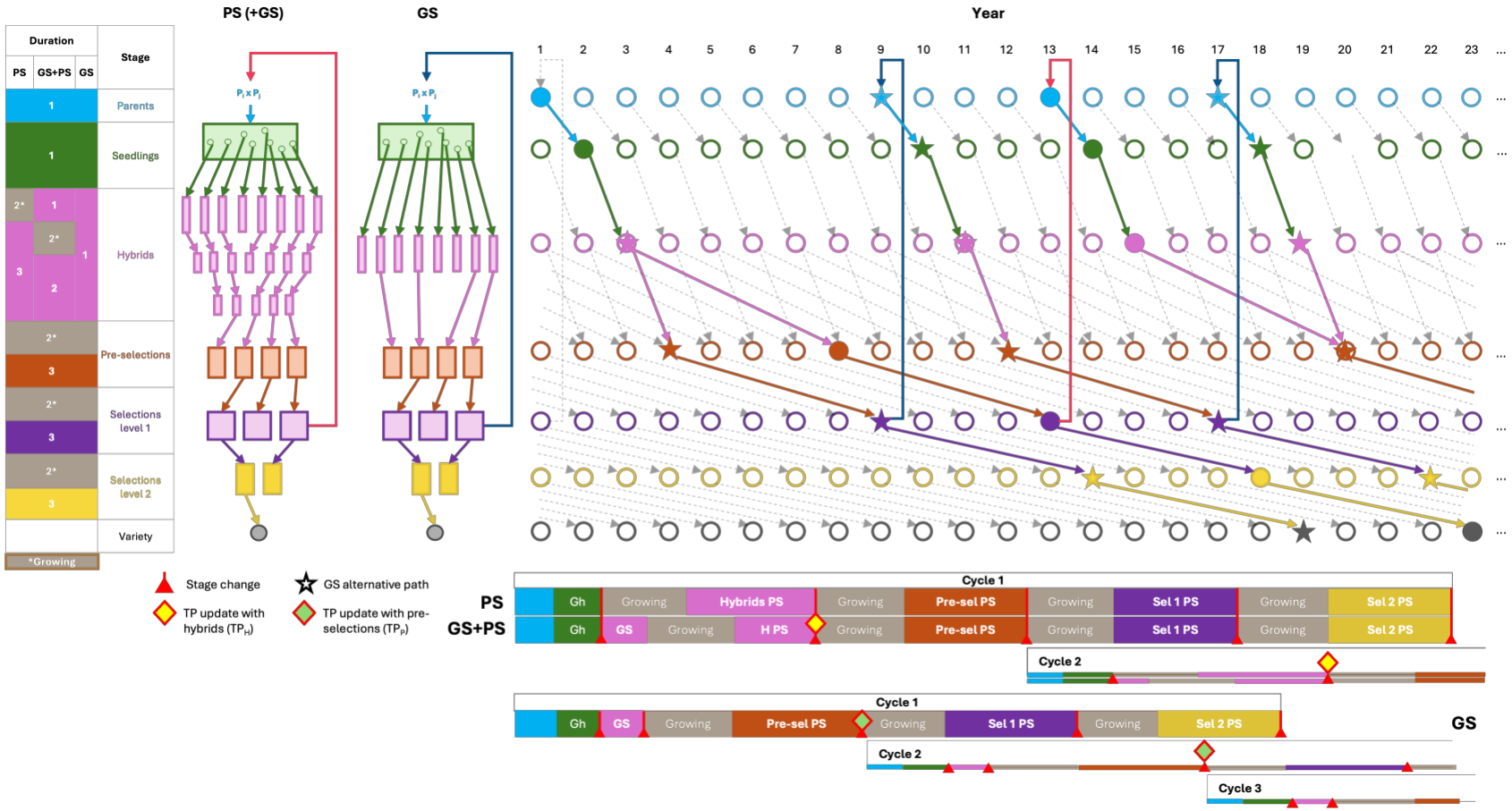


Figure 5. Schematic representation of the conventional PS and alternative GS pathways simulated for the apple (a) and peach (b) breeding program. Diagram showing progression of successive cohorts through breeding stages. Filled circles represent the simulated cohorts while open circles represent possible concurrence of cohorts during the same years in real programs. Arrows show the movement of selected individuals between stages and the reuse of selected candidates as parents. The filled stars represent the alternative GS pathway, highlighting the shortened breeding cycle. Red arrows indicate parental-pool renewal, yellow diamonds indicate training-population updates, and repeated timelines at the bottom illustrate the temporal overlap among consecutive breeding cohorts and the shorter cycle interval enabled by the genomic alternative. TP update occurs only after selected genotypes have both genomic and phenotypic information, which differs between TP_h and TP_p.

One hundred replicates were run for each scenario. The results were aggregated to calculate the statistics, including mean genetic gain, mean genetic variance, mean inbreeding coefficient (F), IBD segment share, LD (r^2), and number of realized segregating sites. IBD segment share, LD and realized segregating sites were reported only for the parental pool during the burn-in to monitor the initialization of the parental pool (fig. supl. 2 and 3).

IV. Results

Genetic gain was simulated for each crop over three breeding cycles. It differed substantially among selection strategies, heritability levels, and species. Nevertheless, some consistent patterns emerged. In both apple and peach, the strategy combining parental mating selection with GPCP, PS, GS, and TP updating with larger hybrid populations (GPCP+PS+GS+TP_H) produced the highest genetic gain at all three heritability levels (fig. 6).

Regarding genetic diversity trends over time, parental pool genetic variance (σ^2_G) followed an expected pattern during the burn-in, with an initial increase, possible instability, and then a decrease and the beginning of stabilization. Then, it began to decline at different rates once selection scenarios started. GPCP schemes, particularly GPCP+GS+TP_P and GPCP+PS+GS+TP_H, showed the fastest decline in variance, while PS retained the most residual genetic variance by the end of the simulated cycles. This pattern was mirrored in the inbreeding coefficient (F) of the parental pool, with PS and the GS+TP_P and PS+GS+TP_H schemes accumulating inbreeding more gradually (fig. 6).

4.1 Genetic gain in apple

In apple, genetic gain increased with heritability under most strategies. The GPCP+PS+GS+TP_H strategy produced gains of 5.02, 5.50, and 5.74 under low, intermediate, and high heritability, respectively. It was therefore the highest-ranking strategy across the full heritability range.

The advantage of GS strategies over PS decreased as heritability increased. At low heritability, GPCP+PS+GS+TP_H gained 5.02, vs. 3.63 for PS, a 38% improvement. At intermediate heritability, gains were 5.50 vs. 4.87, a 13% improvement. At high heritability, PS reached 5.37, close to the 5.74 of the best strategy, reducing the advantage to 7% (Table 3a).

Strategies using GS as the only method of hybrid selection performed poorly in apple. GS+TP_P produced gains of only 3.04-3.35 and ranked last at every heritability level. Similarly, GPCP+GS+TP_P produced lower gains than the strategies that retained phenotypic information. Thus, neither the genomic selection of crosses nor the associated modifications to the breeding scheme, including changes in the number of crosses and the allocation of candidates, were sufficient to compensate for the loss of phenotypic information during subsequent candidate selection.

PS+GS+TP_H strategy also failed to outperform conventional PS. Its genetic gain ranged from 3.50 to 4.44, whereas PS ranged from 3.63 to 5.37, indicating that simply combining phenotypic values and GEBVs during hybrid selection was insufficient.

GS prediction accuracy, measured as the correlation between GEBVs and genetic values (GVs) and between GEBVs and phenotypic values at the hybrid selection stage, showed differences among strategies as breeding cycles progressed. All strategies started from the same initial TP of 399 records and therefore had comparable prediction accuracy at the beginning of the simulation. Differences emerged after TP updating began. Strategies using TP_H generally maintained the highest prediction accuracy across subsequent cycles, with 1,050 hybrid records added to the TP per cycle, compared with 20 pre-selection records per cycle under TP_P.

4.2 Genetic gain in peach

In peach, GPCP+PS+GS+TP_H achieved the highest genetic gain at the end of the three breeding cycles across all heritability levels, reaching 4.16, 4.70, and 4.98. However, when strategies were compared at the same time point, GPCP+GS+TP_P generally showed the highest genetic gain, highlighting its stronger short-term response. More broadly, and unlike in apple, strategies relying on GS alone or in combination with GPCP maintained a substantial advantage over PS across the full range of heritability levels (Table 3b).

At low heritability, the best strategy produced approximately 61% more genetic gain than PS, with gains of 4.16 and 2.59, respectively. Its advantage remained approximately 40% at intermediate heritability and 29% at high heritability. Although the relative benefit decreased as phenotypic accuracy increased, conventional PS did not approach the top-performing genomic strategies as closely as it did in apple (fig. 6).

GPCP strategies that used GS-only also performed better in peach. Under low heritability, GPCP+GS+TP_P produced a gain of 4.13, which was nearly identical to the value of 4.16 obtained by GPCP+PS+GS+TP_H, with the advantage of shortening the breeding process by 12 years if we consider 3 cycles (Table 3, Fig. 6).

As heritability increased, the strategy retaining both genomic and phenotypic hybrid information became clearly superior. At high heritability, GPCP+PS+GS+TP_H produced a gain of 4.98, compared with 4.55 for GPCP+GS+TP_P.

The strategy PS+GS+TP_H produced relatively low gains in peach and did not consistently outperform PS. As in apple, this shows that combining GEBVs and phenotypic values during hybrid selection was not sufficient by itself. Genomic information generated its greatest benefit when it was also used to optimize parental combinations.

Parental pool variance declined after cycle 0 in the same pattern as apple but starting from a lower burn-in plateau, consistent with the smaller effective genome size. GS prediction accuracy was generally higher than in apple.

4.3 Effect of heritability on genetic gain

Overall, higher heritability increased genetic gain, with the effect being especially pronounced for conventional PS. In apple, genetic gain under PS increased from 3.63 at low heritability to 5.37 at high heritability, corresponding to an increase of approximately 48%. In peach, it increased from 2.59 to 3.87, an increase of approximately 49% (fig. 6).

In contrast, integrated PS+GS strategies were less sensitive to changes in heritability. For GPCP+PS+GS+TP_H, genetic gain increased by approximately 14% in apple and 20% in peach between the lowest and highest heritability levels. Consequently, the advantage of GS-based over PS-based strategies was greatest at low heritability and became smaller as heritability increased, although the magnitude of this reduction differed between apple and peach.

4.4 Effect of Optimal Contribution Selection and genetic diversity

Adding OCS generally reduced short-term genetic gain relative to the corresponding GPCP strategy without OCS. For the highest-performing integrated strategy, the reduction associated with OCS ranged from approximately 3% to 6% in apple (fig. 6a) and from

approximately 4% to 7% in peach (fig. 6b). This reduction reflects the trade-off between maximizing short-term genetic gain and preserving genetic diversity over successive breeding cycles.

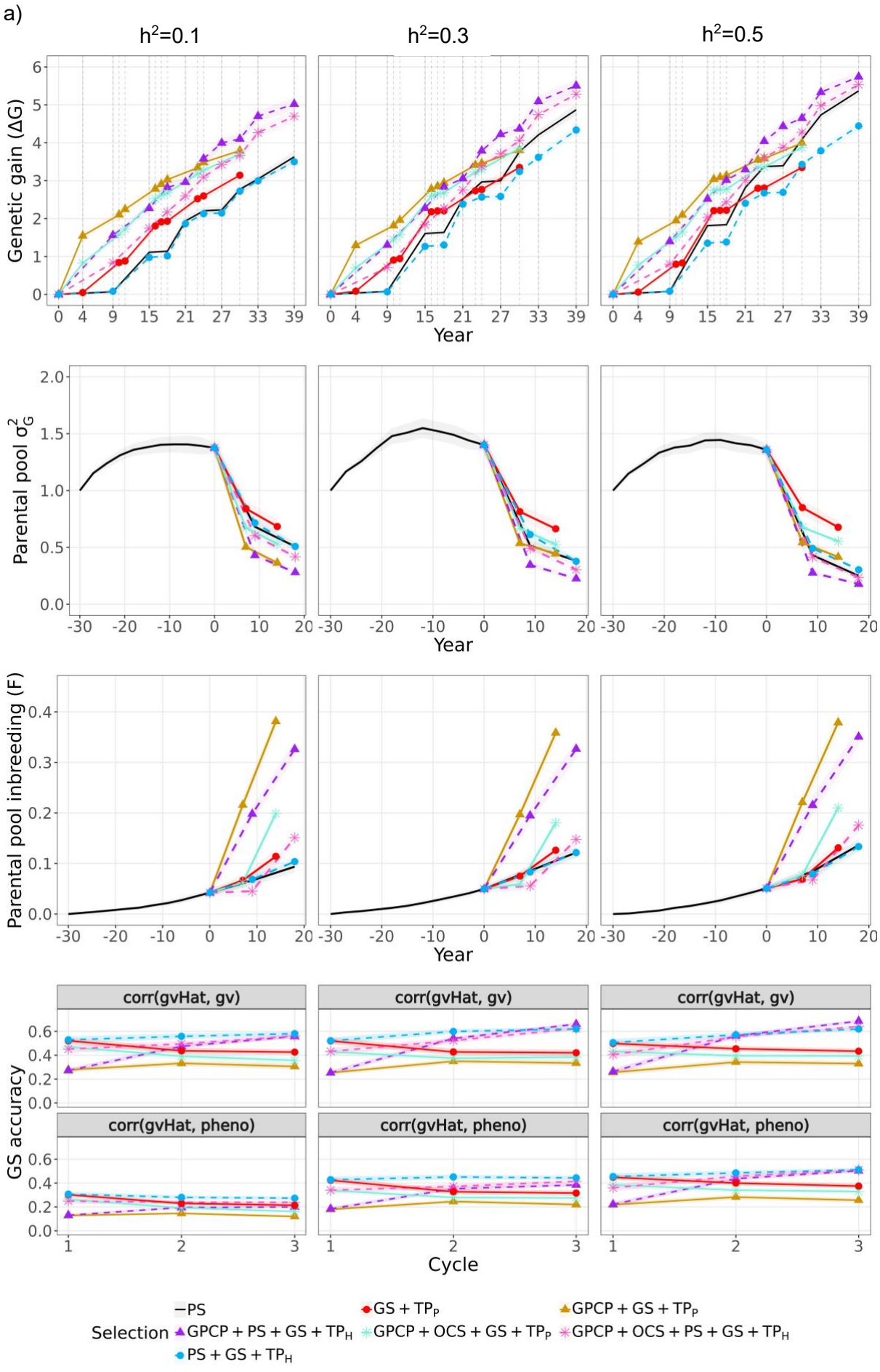
Substantial differences among strategies were also observed for the maintenance of genetic variance. In apple, parental-pool genetic variance decreased from approximately 1.4 at the beginning of the evaluated breeding cycles to values ranging from about 0.2–0.7 by the end of the simulation, depending on the strategy and heritability scenario (fig. 6a). In peach, the corresponding final values ranged approximately from 0.2 to 0.6 (fig. 6b). Importantly, genetic variance preservation was not exclusively associated with the use of OCS. In apple, GS+TP_P and PS+GS+TP_H were among the strategies showing comparatively favorable maintenance of genetic variance.

Differences among strategies were similarly evident for inbreeding. In apple, final parental-pool inbreeding ranged approximately from 0.10 to 0.38, depending on the strategy and heritability level, whereas in peach the difference between the least and most inbred strategies was approximately 0.3 by the end of the simulation.

4.5 The added value of genomics-assisted crosses

The strong performance of the GPCP schemes supports the strategic use of genomic information before crosses are made. Conventional parental selection ranks individual parents, whereas GPCP evaluates the expected performance of their progeny and can therefore exploit complementarity between parents. The results indicate that this use of genomic information at the mating stage provided a substantial additional gain, particularly when combined with subsequent genomic and phenotypic selection.

In apple, adding GPCP to the PS+GS+TP_H strategy increased genetic gain from 3.50 to 5.02 at h^2=0.1, from 4.33 to 5.50 at h^2=0.3, and from 4.44 to 5.74 at h^2=0.5, with relative increases of about 43%, 27%, and 29%. Similarly, in peach, GPCP raised gain from 2.85 to 4.16, from 3.41 to 4.70, and from 3.69 to 4.98 at the same heritability levels, with increases around 46%, 38%, and 35%. Overall, optimizing parental combinations before progeny production in the PS+GS strategy yields substantially more gain than applying genomic and phenotypic selection after crosses. The same pattern was observed when GPCP was combined with GS without the PS component.


a)
h²=0.1
h²=0.3
h²=0.5
Genetic gain (ΔG)
Year
Parental pool σ²G
Year
Parental pool inbreeding (F)
Year
corr(gvHat, gv)
corr(gvHat, pheno)
GS accuracy
Cycle
Selection
PS
GS + TPP
GPCP + GS + TPP
GPCP + PS + GS + TPH
GPCP + OCS + GS + TPP
GPCP + OCS + PS + GS + TPH
PS + GS + TPH

b)

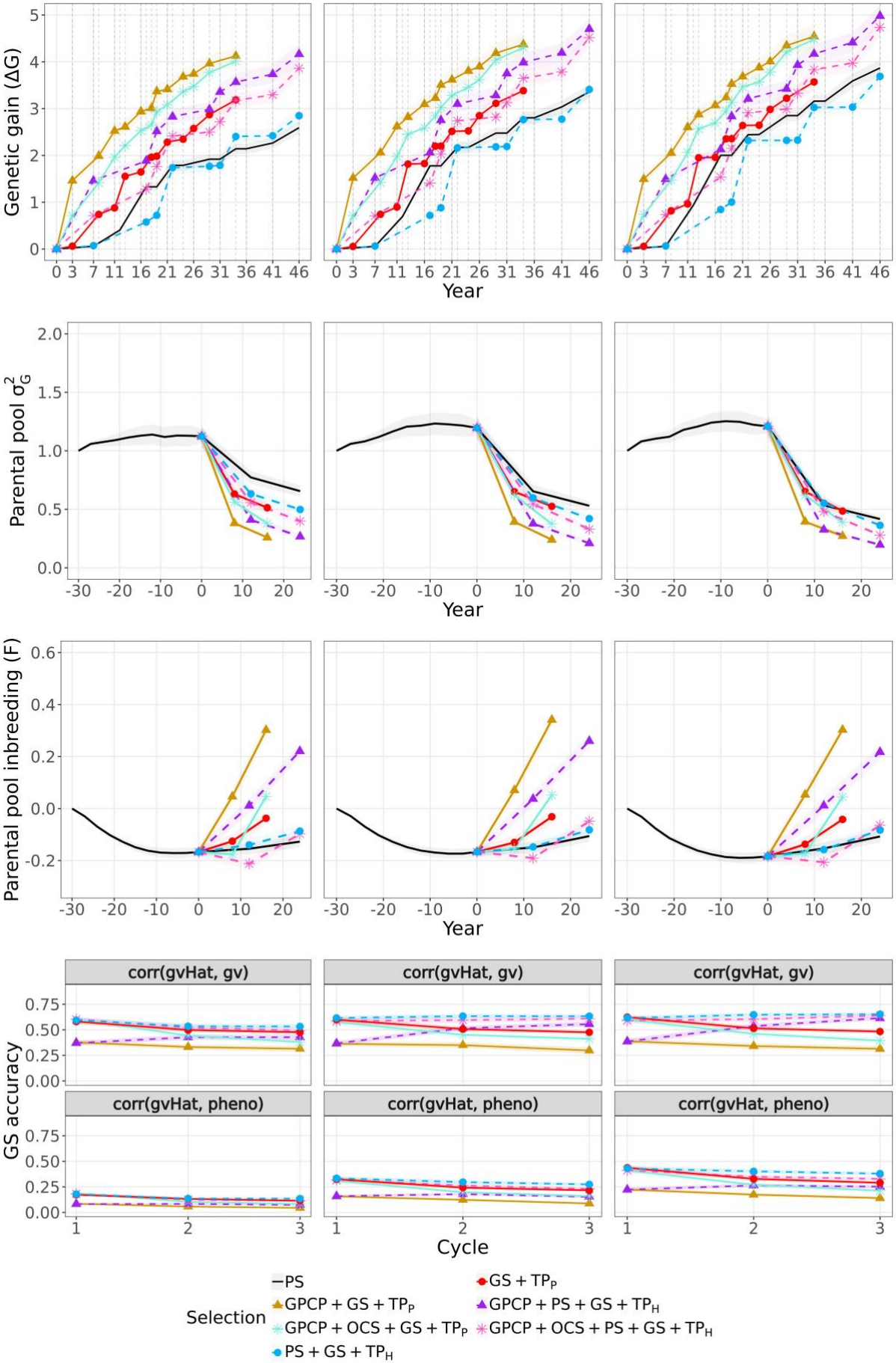

Genetic gain (ΔG)
Year
Parental pool σ^2_G
Year
Parental pool inbreeding (F)
Year
corr(gvHat, gv)
corr(gvHat, pheno)
GS accuracy
Cycle
Selection
PS
GPCP + GS + TP$_P$
GPCP + OCS + GS + TP$_P$
PS + GS + TP$_H$
GS + TP$_P$
GPCP + PS + GS + TP$_H$
GPCP + OCS + PS + GS + TP$_H$

Figure 6. Genetic values, inbreeding coefficients, and selection accuracy of apple (a) and peach (b) alternative GS strategies compared to the PS baseline selection scheme across three levels of heritability.

Table 3. Apple (a) and peach (b) genetic gain results under different selection and mating scheme alternatives and low, medium, or high heritability.

a)

$h^2 = 0.1$		$h^2 = 0.3$		$h^2 = 0.5$	
Selection	**ΔG**	**Selection**	**ΔG**	**Selection**	**ΔG**
GPCP+PS+GS+TP$_H$	5.02	GPCP+PS+GS+TP$_H$	5.50	GPCP+PS+GS+TP$_H$	5.74
GPCP+OCS+PS+GS+TP$_H$	4.70	GPCP+OCS+PS+GS+TP$_H$	5.28	GPCP+OCS+PS+GS+TP$_H$	5.54
GPCP+GS+TP$_P$	3.79	PS	4.87	PS	5.37
GPCP+OCS+GS+TP$_P$	3.70	PS+GS+TP$_H$	4.33	PS+GS+TP$_H$	4.44
PS	3.63	GPCP+OCS+GS+TP$_P$	3.86	GPCP+GS+TP$_P$	3.99
PS+GS+TP$_H$	3.50	GPCP+GS+TP$_P$	3.80	GPCP+OCS+GS+TP$_P$	3.88
GS+TP$_P$	3.04	GS+TP$_P$	3.31	GS+TP$_P$	3.35

b)

$h^2 = 0.1$		$h^2 = 0.3$		$h^2 = 0.5$	
Selection	**ΔG**	**Selection**	**ΔG**	**Selection**	**ΔG**
GPCP+PS+GS+TP$_H$	4.16	GPCP+PS+GS+TP$_H$	4.70	GPCP+PS+GS+TP$_H$	4.98
GPCP+GS+TP$_P$	4.13	GPCP+OCS+PS+GS+TP$_H$	4.52	GPCP+OCS+PS+GS+TP$_H$	4.74
GPCP+OCS+GS+TP$_P$	4.01	GPCP+GS+TP$_P$	4.38	GPCP+GS+TP$_P$	4.55
GPCP+OCS+PS+GS+TP$_H$	3.86	GPCP+OCS+GS+TP$_P$	4.30	GPCP+OCS+GS+TP$_P$	4.48
GS+TP$_P$	3.19	PS+GS+TP$_H$	3.41	PS	3.87
PS+GS+TP$_H$	2.85	GS+TP$_P$	3.38	PS+GS+TP$_H$	3.69
PS	2.59	PS	3.36	GS+TP$_P$	3.57

The value of GPCP was clear when comparing genomic strategies with phenotypic selection. In peach, GPCP+GS+TP$_P$, even without an extra PS step, outperformed PS at every heritability level: 4.13 vs 2.59 at h^2=0.1, 4.38 vs 3.36 at h^2=0.3, and 4.55 vs 3.87 at h^2=0.5, roughly 60%, 30%, and 18% more gain. In apple, GPCP+GS+TP$_P$ only slightly exceeded PS at low heritability (3.79 vs 3.63) and performed worse at h^2 = 0.3 and 0.5 (3.80 vs 4.87 and 3.99 vs 5.37, respectively). This weaker performance in apple may partly reflect the time lag

in updating the training set, since genomic predictions relied on phenotypic information from the previous breeding cycle, a limitation specific to the apple scheme that was not present in peach. Thus, the lower performance of genomic mating alone in apple may result not only from the increasing efficiency of phenotypic selection at higher heritability, but also from reduced prediction relevance caused by delayed training-set updating.

The greatest gains in both species came when GPCP was used as part of an integrated strategy, not just a replacement for later selection. GPCP+PS+GS+TP_H ranked first across all heritability scenarios, with gains of 5.02-5.74 in apple and 4.16-4.98 in peach. Compared to baseline PS, its advantage was largest at low heritability: about 38% in apple (5.02 vs. 3.63) and 61% in peach (4.16 vs. 2.59). The advantage decreased as heritability increased, to about 7% in apple and 29% in peach at h^2=0.5.

Overall, these results identify mating design as an important point of leverage for genomic information. Once the parental set has been chosen, conventional mating cannot recover favorable combinations that were never created. GPCP acts earlier by allocating crosses toward parental combinations expected to generate superior progeny, thereby improving the genetic distribution entering subsequent selection stages.

V. Discussion

This study aimed to assess the added value of genomic selection relative to conventional phenotypic selection under fixed budget and operational constraints, while identifying the stages of the breeding pipeline where genomic information can be used most effectively. By comparing alternative implementations in apple and peach, the analysis also sought to determine how species-specific biological features and trait heritability influence GS efficiency.

These results suggest that, among the many strategies for adopting GS, choices should be driven by where and under which trait architecture it adds the most value. Across species and heritability levels, the combined strategy of genomic parental mating, phenotypic, and genomic hybrid selection (GPCP+PS+GS+TP_H) consistently outperformed conventional PS and GS-only. This indicates that genomic information was most effective when used at multiple decision points and complemented with phenotypic data. With similar budgets within species, the question is therefore not whether GS should replace PS, but where it offers enough accuracy or cycle-time benefits to justify its cost.

Simulation studies in clonally propagated crops have shown that cross-performance prediction can outperform selection based only on parental GEBVs, particularly when total genetic value and non-additive effects are relevant (Werner et al., 2023).

In the present study, however, the simulated trait was purely additive. The advantage of GPCP must therefore have resulted primarily from directing crosses toward higher expected progeny means and allocating the available family sizes more effectively, rather than from exploiting dominance. This distinction matters because the current results show the advantages of a genomic mating design in apple and peach even under an additive model, while the additional value of modeling dominance remains to be tested for these crops.

The comparison between GPCP+PS+GS+TP_H and PS+GS+TP_H indicates that applying GS to hybrid selection alone did not account for the main improvement in genetic gain. Both strategies combined genomic and phenotypic information during early candidate selection and used a larger hybrid-based TP update, but the strategy that also optimized parental crosses consistently ranked first. This suggests that, in these simulations, cross design contributed substantially to the overall response and was at least as important as candidate ranking. By increasing the expected value of the progeny before candidate selection, GPCP allowed genomic information to be used earlier in the breeding process rather than relying only on the subsequent ranking of hybrids. Importantly, this optimization did not add genotyping costs since parental candidates were already genotyped.

Moreover, the weak performance of PS+GS+TP_H alone should not be interpreted as evidence that combining phenotypic and genomic selection is inherently ineffective. In this study, candidates were ranked using the sum of their phenotypic and GEBV ranks, which assigned equal importance to both sources of information regardless of their respective accuracy. The combined strategy also differed from the other schemes in population size, duration, and resource allocation. Alternative selection indices could therefore produce different outcomes. For example, phenotypic and genomic information could be combined using weights proportional to their relative prediction accuracy or reliability, giving greater influence to the source of information that is more robust for a given trait and selection stage.

5.1 Heritability determines the comparison

The relative advantage of genomic strategies decreased as heritability increased. In apple, the advantage of the best scheme over PS declined from 38% to 7% between low and high heritability, while in peach, it declined from 61% to 29%. This pattern is expected because

improving heritability increases the accuracy of direct phenotypic information and therefore strengthens conventional selection. GS is most likely to add value when phenotypes are noisy, expensive, expressed late, or require multi-year and multi-environment evaluation (Heffner et al., 2010; Lorenz, 2013; Wartha and Lorenz, 2021), which is typically the case for many fruit breeding targets such as yield components (flower density, fruit size, etc.) and fruit quality.

This implies that for highly heritable traits in apple breeding programs, the additional genotyping investment required by GS may not be justified on genetic gain alone. In peach, by contrast, GS-containing schemes retained a more consistent advantage across the heritability range tested, suggesting that peach breeders may see a more consistent return on GS investment regardless of trait architecture.

5.2 Optimizing training population to sustain prediction accuracy

Strategies updated with the larger hybrid cohort (TP_H) tended to maintain higher genomic prediction accuracy across cycles than strategies updated only with pre-selections (TP_P). This is consistent with the importance of TP size and representation of the target candidates. Genomic prediction partly captures genetic relationships between the TP and the selection candidates, and accuracy can decline as these populations become more genetically distant (Habier et al., 2007).

Apple studies have similarly shown that prediction accuracy can vary strongly among target families and can improve when the TP is tailored to the candidates or combines complementary sources of elite and genetic-resource material (Muranty et al., 2015; Roth et al., 2020).

The apparent advantage of TP_H should nevertheless be interpreted as the effect of a complete update strategy rather than TP size alone. TP_H differed from TP_P in the number and type of records added, the stage at which they were collected, and the associated phenotyping scheme. The current design therefore cannot isolate whether accuracy was maintained because more records were added, because those records were more closely related to the next candidates, because they better captured intra-family variance, or because they represented a more diverse, less-selected sample. Under the heritability levels tested, the unreplicated design used to phenotype hybrids added to the TP in the TP_H strategy did not penalize genomic prediction accuracy. This may reflect the benefit of allocating phenotyping

resources to a larger number of distinct genotypes rather than to repeated measurements of fewer genotypes, thereby increasing the representation of alleles and genetic combinations in the TP.

5.3 Comparison between crops

An original aspect of this thesis is the simulation and comparison of alternative genomic and phenotypic selection strategies in two perennial fruit crops with contrasting breeding systems. Simulation has previously been used to investigate the placement of GS within plant breeding pipelines, including alternative GS and genomic cross-prediction strategies in clonally propagated crops (Werner et al., 2023). However, the present study extends this approach by applying comparable selection concepts to crop-specific apple and peach breeding programs while explicitly accounting for their respective operational structures and costs. This cross-crop comparison provides an opportunity to assess not only whether GS can increase genetic gain, but also how its relative value changes when implemented within different perennial breeding contexts.

The relative differences in performance of the strategies between apple and peach should be interpreted taking into consideration not only the genome biology but also the specifics of the breeding programs chosen for this study. The apple and peach programs differed in their operational structure, population sizes, selection intensities, replication, unit costs, marker datasets, and TP dynamics. All these factors can influence the relative accuracies and economic feasibility of GS.

GS was more consistently advantageous in the simulated peach program, and prediction accuracies were generally higher than in apple. Biological differences provide plausible hypotheses for this contrast. Cultivated peach has comparatively high co-ancestry and extended shared haplotypes, which can strengthen relationship-based prediction within a breeding population, whereas apple germplasm contains broader diversity and more heterogeneous relationships (Laurens et al., 2018).

Genomic prediction has also shown promising results in peach, although it has been explored much less extensively than in apple. Biscarini et al. (2017), for example, evaluated genome-enabled prediction for fruit weight, sugar content, and titratable acidity in 11 peach progenies and obtained generally moderate to high predictive abilities, although predictions were evaluated within individual progenies rather than across independent families. More recently, multi-environment genomic prediction has been investigated for soluble solids content using

several thousand genome-wide markers (Hardner et al., 2022). In apple, genomic prediction has been studied more extensively across breeding populations and environments, with prediction accuracy shown to depend strongly on the target family, training-population composition, trait architecture, and environment (Muranty et al., 2015; Roth et al., 2020; Jung et al., 2022, reviewed in Muranty et al. 2023).

These biological explanations should be considered hypotheses rather than conclusions because comparisons between species included non-biological differences in breeding-program structure. Indeed, apple and peach breeding steps involved different population sizes, family sizes, selection pressures, stage durations, replication methods, parental-renewal timing, marker datasets, and TP dynamics.

The apple GPCP model also used a TP lagged by one breeding cycle, while the peach model used the TP from the current cycle. This timing difference might have diminished the usefulness of cross prediction for apple. Thus, the results support understanding the relative ranking of schemes within each crop-specific program but do not completely test whether peach biology inherently makes GS more effective than apple biology. Nevertheless, under the current breeding program configurations at NOVADI and CEP, the implementation of GS appears to be more readily facilitated in peach, even though phenotyping costs are relatively low.

The current results also do not demonstrate that WGS is required for apple. Physical genome size alone is not the determinant of genomic prediction accuracy. Marker distribution, LD, genotyping quality, TP relatedness, and the number of independently segregating chromosome segments are more directly relevant (Daetwyler et al., 2008; Goddard, 2009).

A recent apple study found similar predictive ability with 7,255 well-distributed SNPs and more than 300,000 SNPs, while a poorly positioned overlapping marker set performed much worse (Jung et al., 2025). Therefore, increasing marker density or adopting WGS should be evaluated through marker-density and genotyping-cost analyses rather than assumed to improve economic efficiency automatically.

This economic trade-off differed substantially between the two breeding programs considered here, with the ratio of genotyping to phenotyping costs reaching approximately 5.6 in peach compared with 2.2 in apple. Thus, despite the stronger performance of GS in the simulated peach program, genotyping currently represents a comparatively larger cost relative to phenotyping in peach. A substantial reduction in genotyping costs could therefore markedly

improve the cost-effectiveness of GS in both crops and potentially alter the relative attractiveness of genomic strategies.

Overall, the contrasting behavior of GS in the simulated apple and peach programs illustrates the importance of evaluating genomic breeding strategies within the biological and operational context of each crop. Differences in genome architecture, mating system, domestication history, breeding-program structure, and the timing and composition of TP updates may all contribute to the observed responses. Further factorial simulations varying TP size, TP relatedness, phenotypic precision, marker density, and breeding-program structure independently would be required to disentangle these effects and determine which factors most strongly influence the cost-effectiveness of GS across perennial fruit crops.

5.4 Genetic gain and genetic diversity management

The strategies producing the greatest genetic gain also caused the fastest reduction in parental genetic variance and the greatest accumulation of inbreeding. This is an expected consequence of strong recurrent selection, because rapid changes in allele frequency can deplete genetic variance and increase the probability that favorable alleles present at low frequency are lost. Earlier simulations of genomic selection have similarly shown that rapid early gain can be accompanied by loss of favorable alleles, genetic variance, and eventually selection response (Jannink, 2010). Sustained response depends on both short-term selection and maintaining or replacing genetic diversity through effective population size, mutation, or germplasm introduction.

The observed pattern therefore adds an important qualification to the genetic-gain ranking. GPCP+PS+GS+TP_H was the best short-term strategy, but it was not necessarily the most sustainable strategy over a longer breeding perspective. This issue is particularly relevant when GS is used recurrently, because shortened generation intervals can accelerate both genetic gain and the loss of diversity. Long-term GS simulations have consequently emphasized the need to manage gain and diversity jointly rather than considering prediction and selection intensity alone (Allier et al., 2019; Gorjanc et al., 2018; Jannink, 2010).

Adding OCS reduced short-term gain in the GPCP+PS+GS+TP_H strategy by approximately 3-7% across heritability levels. In apple, genetic gain decreased from 5.02 to 4.70, 5.50 to 5.28, and 5.74 to 5.54 at h^2=0.1, 0.3, and 0.5, corresponding to reductions of 6.4%, 4.0%, and 3.5%, respectively. In peach, the corresponding reductions were 7.2%, 3.8%, and 4.8%, with genetic gain decreasing from 4.16 to 3.86, 4.70 to 4.52, and 4.98 to 4.74. Despite this relatively short-

term penalty, OCS also slowed the accumulation of inbreeding and the depletion of parental genetic variance relative to the corresponding unconstrained GPCP strategies (fig. 6).

This short-term penalty is consistent with the objective of OCS, which limits the disproportionate contribution of closely related high-value parents in order to retain diversity for subsequent selection. Longer-term simulations have shown that such constraints can eventually increase cumulative genetic gain. For example, Cowling et al. (2016) simulated 30 cycles of recurrent selection and found that OCS provided more favorable long-term combinations of gain and population coancestry than truncation selection. Gorjanc et al. (2018) simulated recurrent GS over 20 years with one to six selection cycles per year and showed that optimal cross selection improved the efficiency with which genetic diversity was converted into gain and could outperform truncation selection over the longer term. Allier et al. (2019) evaluated 60 annual recurrent-selection cohorts following burn-in and showed that diversity-constrained selection could sacrifice short-term response while producing greater long-term gain.

The simulations covered only a few selection cycles and evaluated OCS at a single target angle of 45°, making them too limited to determine if short-term penalties could be reduced and recovered later. Previous studies show this trade-off depends on the time horizon and diversity constraint. Cowling et al. (2016) compared target angles of 45° and 60° across 30 cycles, while Gorjanc et al. (2018) found the optimal penalty varies with population size and selection cycles. Future evaluation of these strategies should include more cycles and a range of target angles to better understand the balance between immediate response, genetic variance preservation, and long-term gains in perennial breeding.

5.5 Economic interpretation and practical adoption

Because the compared strategies were designed under approximately equivalent total budgets, differences in genetic gain represent differences in breeding efficiency rather than the effect of simply spending more money for GS implementation. Superior genetic gains therefore indicate that reallocating part of the available budget toward genomic mating design and selection could produce greater genetic progress.

This follows the microeconomic framework of Fugeray-Scarbel et al. (2022) and suggests that reallocating resources toward genomic mating design and selective early genotyping can increase genetic progress with the existing budget.

The most important observation points to high PS+GS cost in peach. Although the number of phenotyped candidates was reduced in an attempt to offset the additional genotyping expenses (Table 2b), the relatively low unit cost of phenotyping in the peach program limited the savings that could be achieved through further reductions in field evaluation.

Consequently, decreasing the PS+GS budget to a level comparable with the baseline would likely require reductions in population size or phenotyping intensity that could compromise selection accuracy and the representativeness of the evaluated population. Breeders should therefore consider not only the genetic gain achieved by this strategy but also the magnitude and feasibility of the additional investment required (here 10 k€ to 16k€ compared to alternative PS and GS strategies).

Then, under the simulated conditions for peach, GS-only provides a clearer cost-effectiveness advantage because it achieved genetic gain comparable to or greater than baseline PS across all heritability levels, shortened the breeding-cycle interval, and operated under a budget that was more closely aligned with the conventional scheme. These results suggest that, for peach programs with relatively inexpensive phenotyping, replacing early phenotypic selection with GS may be more economically realistic than adding genomic prediction to an already low-cost phenotypic stage.

However, this comparison is not a complete return-on-investment (ROI) analysis. Total costs were similar but not identical among schemes, and the absolute budgets for apple and peach should not be compared directly because they represent different breeding programs and cost structures. In addition, the simulations assumed that an initial training population (TP) was already available and did not separately account for its establishment or for dedicated recurrent costs associated with model retraining, data curation, computation, and specialist personnel. In contrast, genotyping and phenotyping costs for candidates that could subsequently contribute to TP updating were already included as routine selection operations. Thus, the main unaccounted investments concern establishment of the initial TP and the dedicated analytical infrastructure required to maintain genomic prediction, rather than the routine generation of phenotypic and genotypic data used for TP updating.

For implementation, GPCP appears to be a promising first use of genomic information when a sufficiently accurate and current TP is already available. We did not evaluate GPCP as a stand-alone strategy without GS of progeny. In practice, combining GPCP with GS also creates a recurrent flow of genotyped candidates that are subsequently phenotyped and can be incorporated into model retraining, thereby helping maintain the TP on which GPCP itself

depends. The next priority could therefore be targeted GS, or an optimized PS+GS selection index, for traits and stages where phenotyping is least accurate, most expensive, or available only late in the breeding process.

Nevertheless, in perennial fruit trees, genomic selection cannot eliminate the need for advanced clonal evaluation. Candidate cultivars must still be assessed for several years and environments to validate performance, GxE interaction, fruit and market quality, and ultimately release suitability. Consequently, the breeding cycle cannot be shortened simply to the generation interval achievable through early GS.

However, genomic information could also improve the efficiency of these later testing stages rather than merely replacing early phenotyping. Sparse multi-environment testing provides one such opportunity: because advanced candidates are already genotyped, only a subset of genotype-by-environment combinations may need to be phenotyped, while genomic prediction can be used to infer performance in unobserved combinations and model G×E. In rice, for example, He et al. (2023) found that multi-environment genomic prediction with only about 30% of the phenotypic records could achieve prediction accuracy comparable to substantially higher phenotyping intensities, and that modeling covariance among environments improved prediction (He et al., 2023). Although the reduction in testing for perennial fruit crops is yet to be established, this approach suggests GS could supplement multi-year, multi-location validation by enabling more efficient phenotyping. The phenotypic data from these genotyped selections could be reused in G×E prediction models and future TP updates.

5.6 Limitations and future simulation perspectives

Several assumptions limit the direct transfer of the theoretical gains observed here to commercial breeding. First, the modeled trait was controlled by 200 additive QTLs and excluded dominance, epistasis, GxE interaction, major genes, and correlations with other breeding objectives. Apple and peach breeding programs select multiple traits, often involving trade-offs among fruit quality, productivity, phenology, disease resistance, adaptation, and other targets. Thus, selection responses for one polygenic trait can differ from those under realistic multi-trait goals. Second, MAS was simulated as random selection because the modeled quantitative trait was assumed to be independent of the MAS loci, for example loci controlling disease resistance. The simulations therefore represent selection for one polygenic trait after a neutral bottleneck rather than joint selection on linked or correlated targets. Third,

heritability was imposed through stage-specific error variance, while realized heritability could change across cycles as genetic variance was depleted.

The single-trait framework may also underestimate the relative economic value of genomic information in a real breeding program. A single genotyping event can contribute simultaneously to prediction and selection for several traits, whereas phenotyping additional traits generally requires additional measurements, labor, equipment, or trial resources, even when observations are collected on the same trees. Thus, the marginal cost of adding breeding objectives may increase more rapidly for phenotypic selection than for genomic selection once a common marker dataset and prediction infrastructure are available. Multi-trait simulations could therefore reveal larger relative benefits of GS than those estimated here, although the magnitude of this advantage would depend on genetic correlations among traits, trait-specific prediction accuracy, and the cost structure of phenotyping.

Other limitations concern prediction and breeding operations. The simulations assumed complete marker data, a fixed rrBLUP model, and successful crosses and plant establishment. They did not represent genotyping failures, imputation errors, missing phenotypes, variable seed germination, rootstock-scion interactions, or environmental heterogeneity. In addition, approximate rather than exact equality among breeding budgets and the one-cycle TP lag used for apple GPCP should be considered when interpreting differences among strategies.

More informative next analyses should independently vary genotyping cost, marker density, training-population size and update frequency, family size, parental-renewal timing, and the genetic distance between the training population and selection candidates. Further scenarios should evaluate optimized selection-index weights, dominance and epistasis, G×E architectures, and realistic multi-trait breeding objectives, including favorable and unfavorable genetic correlations among economically important traits. Longer simulations would also be useful for assessing whether short-term gains from intensive genomic selection are maintained when genetic diversity declines and whether diversity-management strategies alter long-term rankings.

These extensions could be implemented by expanding the R Shiny application and simulation pipeline developed in this thesis. The existing framework provides a basis to systematically add biological, operational, and economic parameters and for comparing breeding strategies across a wider range of assumptions. Extending the application in this way would also allow breeders to explore program-specific scenarios interactively, rather than relying on a single set of fixed simulation parameters.

VI. Conclusions

This thesis evaluated whether genomic selection can improve genetic gain in apple and peach breeding using data from real breeding populations and program structures from NOVADI and CEP Innovation. Conventional and genomic strategies were compared under realistic operational constraints and approximately equivalent within-species budgets. The results show that the value of genomic information depends less on its mere inclusion in a breeding program than on the stage at which it is applied.

Across both species and all heritability levels, the highest genetic gain was obtained when genomic prediction was used to optimize parental crosses through GPCP and was subsequently combined with phenotypic and genomic information during hybrid selection. This indicates that genomic tools are most effective as components of an integrated breeding strategy rather than as universal substitutes for phenotypic evaluation.

The relative benefit of genomic strategies also depended on trait heritability and breeding-program structure. Their advantage was more significant for low-heritability traits. As heritability increased, conventional PS became more competitive, particularly in apple. In peach, genomic strategies remained advantageous across a broader heritability range.

From a practical perspective, GPCP is promising for early use of genomic information because it improves the expected value of progeny before establishing a large field. Nevertheless, the preferred implementation may differ between crops. In apple, the integrated GPCP+PS+GS strategy generated the greatest gain, whereas genomic selection without genomic mating optimization produced limited benefits under the simulated conditions. In peach, the strategy combining GPCP with GS, but without the additional PS step during hybrid selection, achieved genetic gain comparable to or greater than baseline PS across the tested heritability levels, shortened the breeding-cycle interval, and maintained a budget closer to the conventional scheme than the more expensive combined GPCP+PS+GS alternative. Thus, the peach results should not be interpreted as evidence that GS alone is sufficient, but rather that genomic selection was most effective when supported by prior genomic optimization of parental crosses.

Overall, GS can improve genetic gain in fruit tree breeding, but its success depends on strategic integration into the existing pipeline. The most robust suggestion from these analyses is to use genomic information where it changes decisions earliest and most effectively, i.e., to

design better crosses, support selection for traits with limited phenotypic accuracy, and shorten the interval between generations. Phenotypic evaluation remains essential for advanced testing and cultivar release. GS could therefore be seen as a tool for directing breeding resources toward the crosses and candidates with the greatest probability of success.

References

Allier, A., Lehermeier, C., Charcosset, A., Moreau, L., & Teyssèdre, S. (2019). Improving Short- and Long-Term Genetic Gain by Accounting for Within-Family Variance in Optimal Cross-Selection. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.01006

Aranzana, M. J., Decroocq, V., Dirlewanger, E., Eduardo, I., Gao, Z. S., Gasic, K., Iezzoni, A., Jung, S., Peace, C., Prieto, H., Tao, R., Verde, I., Abbott, A. G., & Arús, P. (2019). Prunus genetics and applications after de novo genome sequencing: Achievements and prospects. *Horticulture Research*, *6*(1), 58. https://doi.org/10.1038/s41438-019-0140-8

Bančič, J., Greenspoon, P., Gaynor, R. C., & Gorjanc, G. (2025). Plant breeding simulations with AlphaSimR. *Crop Science*, *65*(1), e21312. https://doi.org/10.1002/csc2.21312

Baumgartner, I. O., Kellerhals, M., Costa, F., Dondini, L., Pagliarani, G., Gregori, R., Tartarini, S., Leumann, L., Laurens, F., & Patocchi, A. (2016). Development of SNP-based assays for disease resistance and fruit quality traits in apple (Malus × domestica Borkh.) and validation in breeding pilot studies. *Tree Genetics & Genomes*, *12*(3), 35. https://doi.org/10.1007/s11295-016-0994-y

Beckman, T. G., & Lang, G. A. (2003). Rootstock Breeding for Stone Fruits. *Acta Horticulturae*, (622), 531–551. https://doi.org/10.17660/ActaHortic.2003.622.58

Ben-Sadoun, S., Fugeray-Scarbel, A., Auzanneau, J., Charmet, G., Lemarié, S., & Bouchet, S. (2021). Integration of genomic selection into winter-type bread wheat breeding schemes: A simulation pipeline including economic constraints. *Crop Breeding,Genetics and Genomics*, *3*. https://doi.org/10.20900/cbgg20210008

Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Théron, A., Poncet, C., Micheletti, D., Kerschbamer, E., Di Pierro, E. A., Larger, S., Pindo, M., Van de Weg, E., Davassi, A., Laurens, F., Velasco, R., Durel, C.-E., & Troggio, M. (2016). Development and validation of the Axiom®Apple480K SNP genotyping array. *The Plant Journal*, *86*(1), 62–74. https://doi.org/10.1111/tpj.13145

Biscarini, F., Nazzicari, N., Bink, M., Arús, P., Aranzana, M. J., Verde, I., Micali, S., Pascal, T., Quilot-Turion, B., Lambert, P., Da Silva Linge, C., Pacheco, I., Bassi, D., Stella, A., & Rossini, L. (2017). Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. *BMC Genomics*, *18*(1), 432. https://doi.org/10.1186/s12864-017-3781-8

Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., Kumar, U., Kumar, R. R., Juliana, P., Mishra, D. C., & Kumar, S. (2022). Genomic Selection: A Tool for Accelerating the Efficiency of Molecular Breeding for Development of Climate-Resilient Crops. *Frontiers in Genetics*, *13*, 832153. https://doi.org/10.3389/fgene.2022.832153

Bühlmann-Schütz, S., Hodel, M., Howard, N. P., Lussi, L., & Patocchi, A. (2025). Low input fast-track (LIFT): An approach for fast introgression and stacking of (R-)genes into advanced apple selections. *Planta*, *262*(4), 93. https://doi.org/10.1007/s00425-025-04780-4

Calvert, M., Evers, B., Wang, X., Fritz, A., & Poland, J. (2020). *Breeding Program Optimization for Genomic Selection in Winter Wheat* (p. 2020.10.07.330415). bioRxiv. https://doi.org/10.1101/2020.10.07.330415

Campoy, J. A., Lerigoleur-Balsemin, E., Christmann, H., Beauvieux, R., Girollet, N., Quero-García, J., Dirlewanger, E., & Barreneche, T. (2016). Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of Prunus avium L. landraces and bred cultivars. *BMC Plant Biology*, *16*, 49. https://doi.org/10.1186/s12870-016-0712-9

Cazenave, X. (2022). *Valorisation des ressources génétiques du pommier dans une population d'amélioration élite grâce à la sélection génomique* (Issue 2022ANGE0010) [Theses, Université d'Angers]. https://theses.hal.science/tel-03828587

Cirilli, M., Bassi, D., & Ciacciulli, A. (2016). Sugars in peach fruit: A breeding perspective. *Horticulture Research*, *3*(1), 15067. https://doi.org/10.1038/hortres.2015.67

Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57–65. https://doi.org/10.1016/j.tig.2013.10.002

Cowling, W. A., Castro-Urrea, F. A., Stefanova, K. T., Li, L., Banks, R. G., Saradadevi, R., Sass, O., Kinghorn, B. P., & Siddique, K. H. M. (2023). Optimal Contribution Selection Improves the Rate of Genetic Gain in Grain Yield and Yield Stability in Spring Canola in Australia and Canada. *Plants*, *12*(2), 383. https://doi.org/10.3390/plants12020383

Cowling, W. A., Li, L., Siddique, K. H. M., Henryon, M., Berg, P., Banks, R. G., & Kinghorn, B. P. (2016). Evolving gene banks: Improving diverse populations of crop and exotic germplasm with optimal contribution selection. *Journal of Experimental Botany*, erw406. https://doi.org/10.1093/jxb/erw406

Da Silva Linge, C., Cai, L., Fu, W., Clark, J., Worthington, M., Rawandoozi, Z., Byrne, D. H., & Gasic, K. (2021). Multi-Locus Genome-Wide Association Studies Reveal Fruit Quality Hotspots in Peach Genome. *Frontiers in Plant Science*, *12*, 644799. https://doi.org/10.3389/fpls.2021.644799

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., … Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, *49*(7), 1099–1106. https://doi.org/10.1038/ng.3886

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. *PLOS ONE*, *3*(10), e3395. https://doi.org/10.1371/journal.pone.0003395

Dardick, C., & Callahan, A. M. (2014). Evolution of the fruit endocarp: Molecular mechanisms underlying adaptations in seed protection and dispersal strategies. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00284

Denancé, C., Muranty, H., & Durel, C.-E. (2021). *FruitBreedomics apple high-density SNP genotyping dataset* [Dataset]. Recherche Data Gouv. https://doi.org/10.15454/IOPGYF

Drazeta, L., Lang, S., Hall, A., Volz, R., & Jameson, P. E. (1999). *Seed set and the development of fruit shape in apple.* (K. M. P. M. T. McManus H. A. Outred, Ed.; Issue No. 12, pp. 99–101). Agronomy Society of New Zealand.

FAOSTAT. (2024). https://www.fao.org/faostat/en/#rankings/countries_by_commodity

Frett, T. J., Gasic, K., Clark, J. R., Byrne, D., Gradziel, T., & Crisosto, C. (2012). Standardized Phenotyping for Fruit Quality in Peach[Prunus Persica(L.) Batsch]. *Journal of the American Pomological Society*, *66*(4), 214–219. https://doi.org/10.71318/apom.2012.66.4.214

Fugeray-Scarbel, A., Ben-Sadoun, S., Bouchet, S., & Lemarié, S. (2022). Analyzing the Economic Effectiveness of Genomic Selection Relative to Conventional Breeding Approaches. In N. Ahmadi & J. Bartholomé (Eds.), *Genomic Prediction of Complex Traits: Methods and Protocols* (pp. 619–644). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_22

Gasic, K., Reighard, G., Okie, W., Clark, J., Gradziel, T., Byrne, D., Peace, C., Stegmeir, T., Rosyara, U., & Iezzoni, A. (2015). BACTERIAL SPOT RESISTANCE IN PEACH: FUNCTIONAL ALLELE DISTRIBUTION IN BREEDING GERMPLASM. *Acta Horticulturae*, (1084), 69–74. https://doi.org/10.17660/ActaHortic.2015.1084.7

Gaynor, C. (2025). *Traits in AlphaSimR*. https://cran.r-project.org/web/packages/AlphaSimR/vignettes/traits.pdf

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., Mackay, I. J., & Hickey, J. M. (2017). A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. *Crop Science*, *57*(5), 2372–2386. https://doi.org/10.2135/cropsci2016.09.0742

Gaynor, R. C., Gorjanc, G., & Hickey, J. M. (2021). AlphaSimR: An R package for breeding program simulations. *G3 Genes|Genomes|Genetics*, *11*(2), jkaa017. https://doi.org/10.1093/g3journal/jkaa017

GenoToul Bioinfo. (2018). *GenoToul Bioinformatics Facility*. INRAE. https://doi.org/10.15454/1.5572369328961167E12

Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, *136*(2), 245–257. https://doi.org/10.1007/s10709-008-9308-0

Gorjanc, G., Battagin, M., Dumasy, J.-F., Antolin, R., Gaynor, R. C., & Hickey, J. M. (2017). Prospects for Cost-Effective Genomic Selection via Accurate Within-Family Imputation. *Crop Science*, *57*(1), 216–228. https://doi.org/10.2135/cropsci2016.06.0526

Gorjanc, G., Gaynor, R. C., & Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theoretical and Applied Genetics*, *131*(9), 1953–1966. https://doi.org/10.1007/s00122-018-3125-3

Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2007). The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics*, *177*(4), 2389–2397. https://doi.org/10.1534/genetics.107.081190

Hassanpour, A., Geibel, J., Simianer, H., & Pook, T. (2023). Optimization of breeding program design through stochastic simulation with kernel regression. *G3 Genes|Genomes|Genetics*, *13*(12), jkad217. https://doi.org/10.1093/g3journal/jkad217

He, S., Liang, S., Meng, L., Cao, L., & Ye, G. (2023). Sparse Phenotyping and Haplotype-Based Models for Genomic Prediction in Rice. *Rice*, *16*(1), 27. https://doi.org/10.1186/s12284-023-00643-2

Heffner, E. L., Lorenz, A. J., Jannink, J., & Sorrells, M. E. (2010). Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science*, *50*(5), 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

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

Howard, N. P., Troggio, M., Durel, C.-E., Muranty, H., Denancé, C., Bianco, L., Tillman, J., & Van De Weg, E. (2021). Integration of Infinium and Axiom SNP array data in the outcrossing species Malus × domestica and causes for seemingly incompatible calls. *BMC Genomics*, *22*(1), 246. https://doi.org/10.1186/s12864-021-07565-7

Iezzoni, A. F., McFerson, J., Luby, J., Gasic, K., Whitaker, V., Bassil, N., Yue, C., Gallardo, K., McCracken, V., Coe, M., Hardner, C., Zurn, J. D., Hokanson, S., Van De Weg, E., Jung, S., Main, D., Da Silva Linge, C., Vanderzande, S., Davis, T. M., … Peace, C. (2020). RosBREED: Bridging the chasm between discovery and application to enable DNA-informed breeding in rosaceous crops. *Horticulture Research*, *7*(1), 177. https://doi.org/10.1038/s41438-020-00398-7

Iezzoni, A., Weebadde, C., Peace, C., Main, D., Bassil, N. V., Coe, M., Fazio, G., Gallardo, K., Gasic, K., Luby, J., McFerson, J., Van De Weg, E., & Yue, C. (2016). Where are we now as we merge genomics into plant breeding and what are our limitations? Experiences from RosBREED. *Acta Horticulturae*, (1117), 1–6. https://doi.org/10.17660/ActaHortic.2016.1117.1

IRSC 16K SNP array for Prunus persica | GDR. (n.d.). Retrieved July 20, 2026, from https://www.rosaceae.org/Analysis/431

Iwata, H., Minamikawa, M. F., Kajiya-Kanegae, H., Ishimori, M., & Hayashi, T. (2016). Genomics-assisted breeding in fruit trees. *Breeding Science*, *66*(1), 100–115. https://doi.org/10.1270/jsbbs.66.100

Jannink, J.-L. (2010). Dynamics of long-term genomic selection. *Genetics Selection Evolution*, *42*(1), 35. https://doi.org/10.1186/1297-9686-42-35

Jung, M., Hodel, M., Knauf, A., Kupper, D., Neuditschko, M., Bühlmann-Schütz, S., Studer, B., Patocchi, A., & Broggini, G. A. (2025). Evaluation of genomic and phenomic prediction for application in apple breeding. *BMC Plant Biology*, *25*(1), 103. https://doi.org/10.1186/s12870-025-06104-w

Jung, M., Keller, B., Roth, M., Aranzana, M. J., Auwerkerken, A., Guerra, W., Al-Rifaï, M., Lewandowski, M., Sanin, N., Rymenants, M., Didelot, F., Dujak, C., Font i Forcada, C., Knauf, A., Laurens, F., Studer, B., Muranty, H., & Patocchi, A. (2022). Genetic architecture and genomic predictive ability of apple quantitative traits across environments. *Horticulture Research*, *9*, uhac028. https://doi.org/10.1093/hr/uhac028

Jung, M., Roth, M., Aranzana, M. J., Auwerkerken, A., Bink, M., Denancé, C., Dujak, C., Durel, C.-E., Font i Forcada, C., Cantin, C. M., Guerra, W., Howard, N. P., Keller, B., Lewandowski, M., Ordidge, M., Rymenants, M., Sanin, N., Studer, B., Zurawicz, E., … Muranty, H. (2020). The apple REFPOP—a reference population for genomics-assisted breeding in apple. *Horticulture Research*, *7*(1), 189. https://doi.org/10.1038/s41438-020-00408-8

Kinghorn, B. P. (2011). An algorithm for efficient constrained mate selection. *Genetics Selection Evolution*, *43*(1), 4. https://doi.org/10.1186/1297-9686-43-4

Kumar, S., Chagné, D., Bink, M. C. A. M., Volz, R. K., Whitworth, C., & Carlisle, C. (2012). Genomic Selection for Fruit Quality Traits in Apple (Malus×domestica Borkh.). *PLoS ONE*, *7*(5), e36674. https://doi.org/10.1371/journal.pone.0036674

Kumar, S., Hilario, E., Deng, C. H., & Molloy, C. (2020). Turbocharging introgression breeding of perennial fruit crops: A case study on apple. *Horticulture Research*, *7*(1), 47. https://doi.org/10.1038/s41438-020-0270-z

Kumar, S., Volz, R., & Weskett, R. (2011). Genetic architecture of fruit quality traits in Malus x domestica (Borkh.) compared between own-rooted seedlings and vegetative propagules on 'M. 9' rootstock. *Tree Genetics & Genomes*, *7*(5), 1079–1088. https://doi.org/10.1007/s11295-011-0396-0

Laurens, F., Aranzana, M. J., Arus, P., Bassi, D., Bink, M., Bonany, J., Caprera, A., Corelli-Grappadelli, L., Costes, E., Durel, C.-E., Mauroux, J.-B., Muranty, H., Nazzicari, N., Pascal, T., Patocchi, A., Peil, A., Quilot-Turion, B., Rossini, L., Stella, A., … Van De Weg, E. (2018). An integrated approach for increasing breeding efficiency in apple and peach in Europe. *Horticulture Research*, *5*(1), 11. https://doi.org/10.1038/s41438-018-0016-3

Li, X., Wang, J., Su, M., Zhang, M., Hu, Y., Du, J., Zhou, H., Yang, X., Zhang, X., Jia, H., Gao, Z., & Ye, Z. (2023). Multiple-statistical genome-wide association analysis and genomic prediction of fruit aroma and agronomic traits in peaches. *Horticulture Research*, *10*(7), uhad117. https://doi.org/10.1093/hr/uhad117

Linge, C. da S., Antanaviciute, L., Abdelghafar, A., Arús, P., Bassi, D., Rossini, L., Ficklin, S., & Gasic, K. (2018). High-density multi-population consensus genetic linkage map for peach. *PLOS ONE*, *13*(11), e0207724. https://doi.org/10.1371/journal.pone.0207724

Lirong, W., Yong, L., Gengrui, Z., Weichao, F., Changwen, C., Ke, C., & Xinwei, W. (2020). Peach genomics and breeding programs at Zhengzhou Fruit Research Institute, CAAS. *Acta Horticulturae*, (1282), 1–6. https://doi.org/10.17660/ActaHortic.2020.1282.1

Liu, H., Tessema, B. B., Jensen, J., Cericola, F., Andersen, J. R., & Sørensen, A. C. (2019). ADAM-Plant: A Software for Stochastic Simulations of Plant Breeding From Molecular to Phenotypic Level and From Simple Selection to Complex Speed Breeding Programs. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.01926

Lorenz, A. J. (2013). Resource Allocation for Maximizing Prediction Accuracy and Genetic Gain of Genomic Selection in Plant Breeding: A Simulation Experiment. *G3 Genes|Genomes|Genetics*, *3*(3), 481–491. https://doi.org/10.1534/g3.112.004911

Lush, J. L. (1937). *Animal breeding plans.* (p. viii + [2] + 350 pp.). Collegiate Press Inc.

Magon, G., De Rosa, V., Martina, M., Falchi, R., Acquadro, A., Barcaccia, G., Portis, E., Vannozzi, A., & De Paoli, E. (2023). Boosting grapevine breeding for climate-smart viticulture: From genetic resources to predictive genomics. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1293186

Malus x domestica | *GDR*. (n.d.). Retrieved April 15, 2026, from https://www.rosaceae.org/organism/Malus/x-domestica

Mauro, R. P., Pérez-Alfocea, F., Cookson, S. J., Ollat, N., & Vitale, A. (2022). Editorial: Physiological and Molecular Aspects of Plant Rootstock-Scion Interactions. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.852518

Migicovsky, Z., Gardner, K. M., Richards, C., Thomas Chao, C., Schwaninger, H. R., Fazio, G., Zhong, G.-Y., & Myles, S. (2021). Genomic consequences of apple improvement. *Horticulture Research*, *8*, 9. https://doi.org/10.1038/s41438-020-00441-7

Miller, T., Fleming, M., Miller, S. B., Saski, C., & Gasic, K. (2019, July 23). *Ppe-Xap a DNA Test for Routine Prediction in Breeding of Peach Bacterial Spot Fruit Resistance*. 2019

ASHS Annual Conference. https://ashs.confex.com/ashs/2019/meetingapp.cgi/Paper/31228

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). *Sustainable data analysis with Snakemake*. F1000Research. https://doi.org/10.12688/f1000research.29032.1

Monet, R., Guye, A., Roy, M., & Dachary, N. (1996). Peach mendelian genetics: A short review and new results. *Agronomie*, *16*(5), 321–329. https://doi.org/10.1051/agro:19960505

Muranty, H., Denancé, C., Feugey, L., Crépin, J.-L., Barbier, Y., Tartarini, S., Ordidge, M., Troggio, M., Lateur, M., Nybom, H., Paprstein, F., Laurens, F., & Durel, C.-E. (2020). Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biology*, *20*(1), 2. https://doi.org/10.1186/s12870-019-2171-6

Muranty, H., Troggio, M., Sadok, I. B., Rifaï, M. A., Auwerkerken, A., Banchi, E., Velasco, R., Stevanato, P., Van De Weg, W. E., Di Guardo, M., Kumar, S., Laurens, F., & Bink, M. C. A. M. (2015). Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research*, *2*(1), 15060. https://doi.org/10.1038/hortres.2015.60

Nyaga, C., Labroo, M. R., Paterne, A., Asfaw, A., Wolfe, M. D., Tecle, I. Y., & Mueller, L. A. (2025). Genomic Predicted cross performance: A tool for optimizing parental combinations in breeding programs. *Database*, *2025*, baaf074. https://doi.org/10.1093/database/baaf074

OConnor, K. M., Hayes, B. J., Hardner, C. M., Alam, M., Henry, R. J., & Topp, B. L. (2021). Genomic selection and genetic gain for nut yield in an Australian macadamia breeding population. *BMC Genomics*, *22*, 370. https://doi.org/10.1186/s12864-021-07694-z

Osorio, M., Fiol, A., Ballesta, P., Ahumada, S., Marambio, P., Martínez-Carrasco, P., Infante, R., & Pacheco, I. (2026). Multi-Season Genome-Wide Association Study Reveals Loci and Candidate Genes for Fruit Quality and Maturity Traits in Peach. *Plants*, *15*(2). https://doi.org/10.3390/plants15020189

Pacheco-Ruiz, P., Osorio, S., & Vallarino, J. G. (2025). From data to decisions: A paradigm shift in fruit agriculture through the integration of multi-omics, modern phenotyping, and cutting-edge bioinformatic tools. *Frontiers in Plant Science*, *16*. https://doi.org/10.3389/fpls.2025.1707289

Peil, A., Kellerhals, M., Höfer, M., & Flachowsky, H. (2011). Apple Breeding – From the Origin to Genetic Engineering. *Methods in Temperate Fruit Breeding*, (1).

Peixoto, M. A., Coelho, I. F., Leach, K. A., Bhering, L. L., & Resende Jr., M. F. R. (2024). Simulation-based decision-making and implementation of tools in hybrid crop breeding pipelines. *Crop Science*, *64*(1), 110–125. https://doi.org/10.1002/csc2.21139

Prunus persica | *GDR*. (n.d.). Retrieved April 15, 2026, from https://www.rosaceae.org/organism/24333

R2D2 Consortium, Fugeray-Scarbel, A., Bastien, C., Dupont-Nivet, M., & Lemarié, S. (2021). Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.629737

Riedelsheimer, C., & Melchinger, A. E. (2013). Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theoretical and Applied Genetics*, *126*(11), 2835–2848. https://doi.org/10.1007/s00122-013-2175-9

Robinson, H., Robles-Zazueta, C. A., & Voss-Fels, K. P. (2025). Accelerating perennial crop improvement via multi-omics-based predictive breeding. *The Plant Genome*, *18*(4), e70058. https://doi.org/10.1002/tpg2.70058

Roth, M., Muranty, H., Di Guardo, M., Guerra, W., Patocchi, A., & Costa, F. (2020). Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. *Horticulture Research*, *7*, 148. https://doi.org/10.1038/s41438-020-00370-5

Roth, M., Serrie, M., Dlalah, N., Signoret, V., Viret, S., Gilles, F., Montrognon, Y., & Quilot-Turion, B. (2022). Genetic diversity in a new peach core-collection designed for resilience breeding. *Acta Horticulturae*, (1352), 141–148. https://doi.org/10.17660/ActaHortic.2022.1352.19

Serrie, M., Ribeyre, F., Brun, L., Audergon, J.-M., Quilot, B., & Roth, M. (2024). Dare to be resilient: The key to future pesticide-free orchards? *Journal of Experimental Botany*, *75*(13), 3835–3848. https://doi.org/10.1093/jxb/erae150

Seyum, E. G., Bille, N. H., Abtew, W. G., Munyengwa, N., Bell, J. M., & Cros, D. (2022). Genomic selection in tropical perennial crops and plantation trees: A review. *Molecular Breeding : New Strategies in Plant Improvement*, *42*(10), 58. https://doi.org/10.1007/s11032-022-01326-4

Shimizu, T., & Nonaka, K. (2025). Genomics unlocks the potential of genetic resources for citrus breeding. *Breeding Science*, *75*(1), 21–33. https://doi.org/10.1270/jsbbs.24047

Tegtmeier, R., Švara, A., Gritsenko, D., & Khan, A. (2025). Malus sieversii: A historical, genetic, and conservational perspective of the primary progenitor species of domesticated apples. *Horticulture Research*, *12*(1), uhae244. https://doi.org/10.1093/hr/uhae244

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Poncet, C., Lateur, M., Houben, P., Ordidge, M., Fernandez-Fernandez, F., Evans, K. M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Miranda, C., Gassmann, J., Kellerhals, M., … Durel, C.-E. (2016). Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology*, *16*(1), 130. https://doi.org/10.1186/s12870-016-0818-0

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S. K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., … Viola, R. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nature Genetics*, *42*(10), 833–839. https://doi.org/10.1038/ng.654

Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M. T., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L., Jenkins, J., Vendramin, E., Meisel, L. A., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., … Rokhsar, D. S. (2013). The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, *45*(5), 487–494. https://doi.org/10.1038/ng.2586

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., Micheletti, D., Rosyara, U. R., Cattonaro, F., Vendramin, E., Main, D., Aramini, V., Blas, A. L., Mockler, T. C., Bryant, D. W., Wilhelm, L., Troggio, M., Sosinski, B., Aranzana, M. J., … Peace, C. (2012). Development and Evaluation of a 9K SNP Array for Peach by Internationally Coordinated SNP Detection and Validation in Breeding Germplasm. *PLOS ONE*, *7*(4), e35668. https://doi.org/10.1371/journal.pone.0035668

Vieira, R. A., Nogueira, A. P. O., & Fritsche-Neto, R. (2025). Optimizing the selection of quantitative traits in plant breeding using simulation. *Frontiers in Plant Science*, *16*. https://doi.org/10.3389/fpls.2025.1495662

Villiers, K., Voss-Fels, K. P., Dinglasan, E., Jacobs, B., Hickey, L., & Hayes, B. J. (2024). Evolutionary computing to assemble standing genetic diversity and achieve long-term genetic gain. *The Plant Genome*, *17*(2), e20467. https://doi.org/10.1002/tpg2.20467

Wartha, C. A., & Lorenz, A. J. (2021). Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities. *Crop Breeding and Applied Biotechnology*, *21*, e394621S15. https://doi.org/10.1590/1984-70332021v21Sa28

Way, R. D., Aldwinckle, H. S., Lamb, R. C., Rejman, A., Sansavini, S., Shen, T., Watkins, R., Westwood, M. N., & Yoshida, Y. (1991). APPLES (MALUS). *Acta Horticulturae*, (290), 3–46. https://doi.org/10.17660/ActaHortic.1991.290.1

Werner, C. R., Gaynor, R. C., Sargent, D. J., Lillo, A., Gorjanc, G., & Hickey, J. M. (2023). Genomic selection strategies for clonally propagated crops. *Theoretical and Applied Genetics*, *136*(4), 74. https://doi.org/10.1007/s00122-023-04300-6

Yamamoto, T., & Terakami, S. (2016). Genomics of pear and other Rosaceae fruit trees. *Breeding Science*, *66*(1), 148–159. https://doi.org/10.1270/jsbbs.66.148

Yu, W., & Cheng, Q. (2024). Accelerating Yam Breeding Cycles with Genomic Selection. *Genomics and Applied Biology*. https://doi.org/10.5376/gab.2024.15.0033

Yu, Y., Fu, J., Xu, Y., Zhang, J., Ren, F., Zhao, H., Tian, S., Guo, W., Tu, X., Zhao, J., Jiang, D., Zhao, J., Wu, W., Wang, G., Ma, R., Jiang, Q., Wei, J., & Xie, H. (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nature Communications*, *9*(1), 5404. https://doi.org/10.1038/s41467-018-07744-3

Yue, C., Gallardo, R. K., Luby, J., Rihn, A., McFerson, J. R., McCracken, V., Bedford, D., Brown, S., Evans, K., Weebadde, C., Sebolt, A., & Iezzoni, A. F. (2013). An Investigation of U.S. Apple Producers' Trait Prioritization—Evidence from Audience Surveys. *HortScience*, *48*(11), 1378–1384. https://doi.org/10.21273/HORTSCI.48.11.1378

Zhang, C., Serra, S., Quirós-Vargas, J., Sangjan, W., Musacchi, S., & Sankaran, S. (2023). Non-invasive sensing techniques to phenotype multiple apple tree architectures. *Information Processing in Agriculture*, *10*(1), 136–147. https://doi.org/10.1016/j.inpa.2021.02.001

Zhang, X., Tang, M., Li, J., Chi, Y., Wang, K., Peng, J., & Zhang, Y. (2025). Phenotypic Characters and Inheritance Tendency of Agronomic Traits in F1 Progeny of Pear. *Plants*, *14*(10). https://doi.org/10.3390/plants14101491

Zheng, Y., Crawford, G. W., & Chen, X. (2014). Archaeological Evidence for Peach (Prunus persica) Cultivation and Domestication in China. *PLoS ONE*, *9*(9), e106595. https://doi.org/10.1371/journal.pone.0106595

Supplementary material

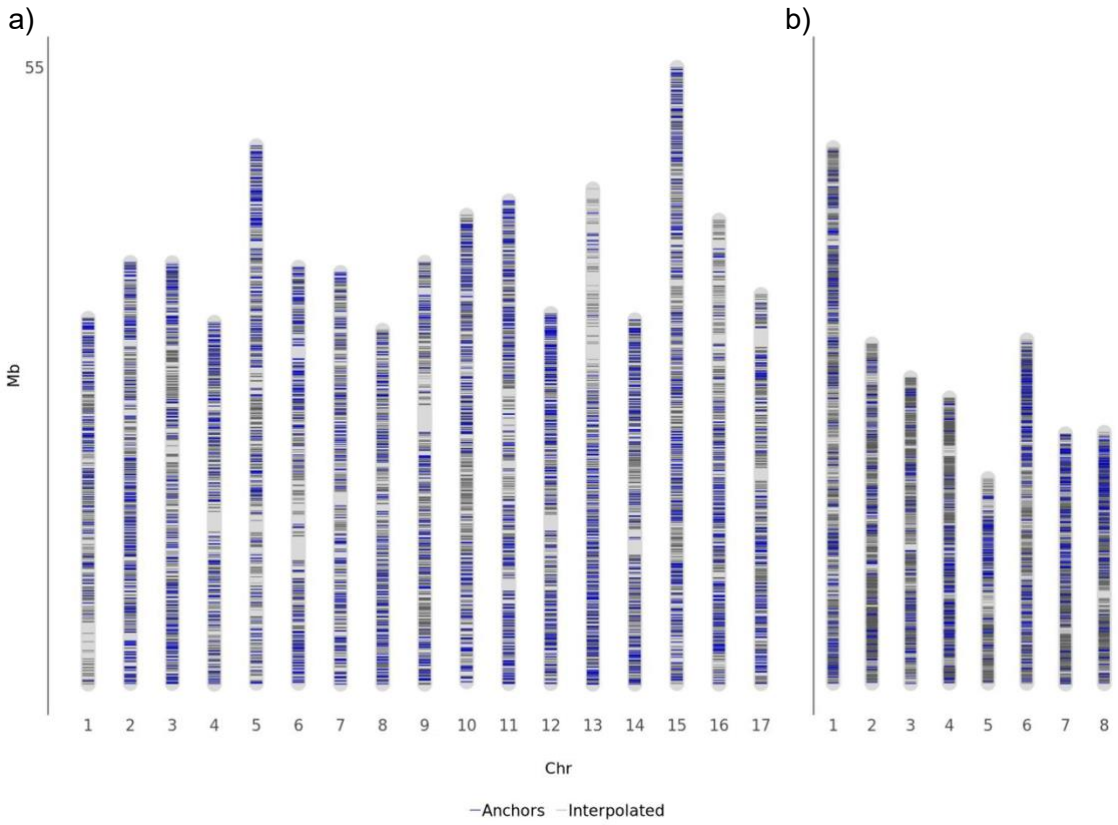


Figure 1. Segregant sites across apple (a) and peach (b) chromosomes. SNP markers from th Axiom® Apple480K array dataset and IRSC 16k Illumina SNP array for *Prunus persica* are used to model segregant sites assigned to markers and QTLs associated with the simulated trait. Monotonic interpolation was used to infer the genetic position of markers with physical position only. For apple: 24,172 marker positions (Mb) in the marker matrix and 10,290 marker positions (cM) in the consensus genetic map, from which 5,146 positions were fixed as anchors for the interpolation. For peach: 13,683 segregating marker positions (Mb) and 3,092 marker positions in the consensus genetic map, from which 2,930 were anchor positions.

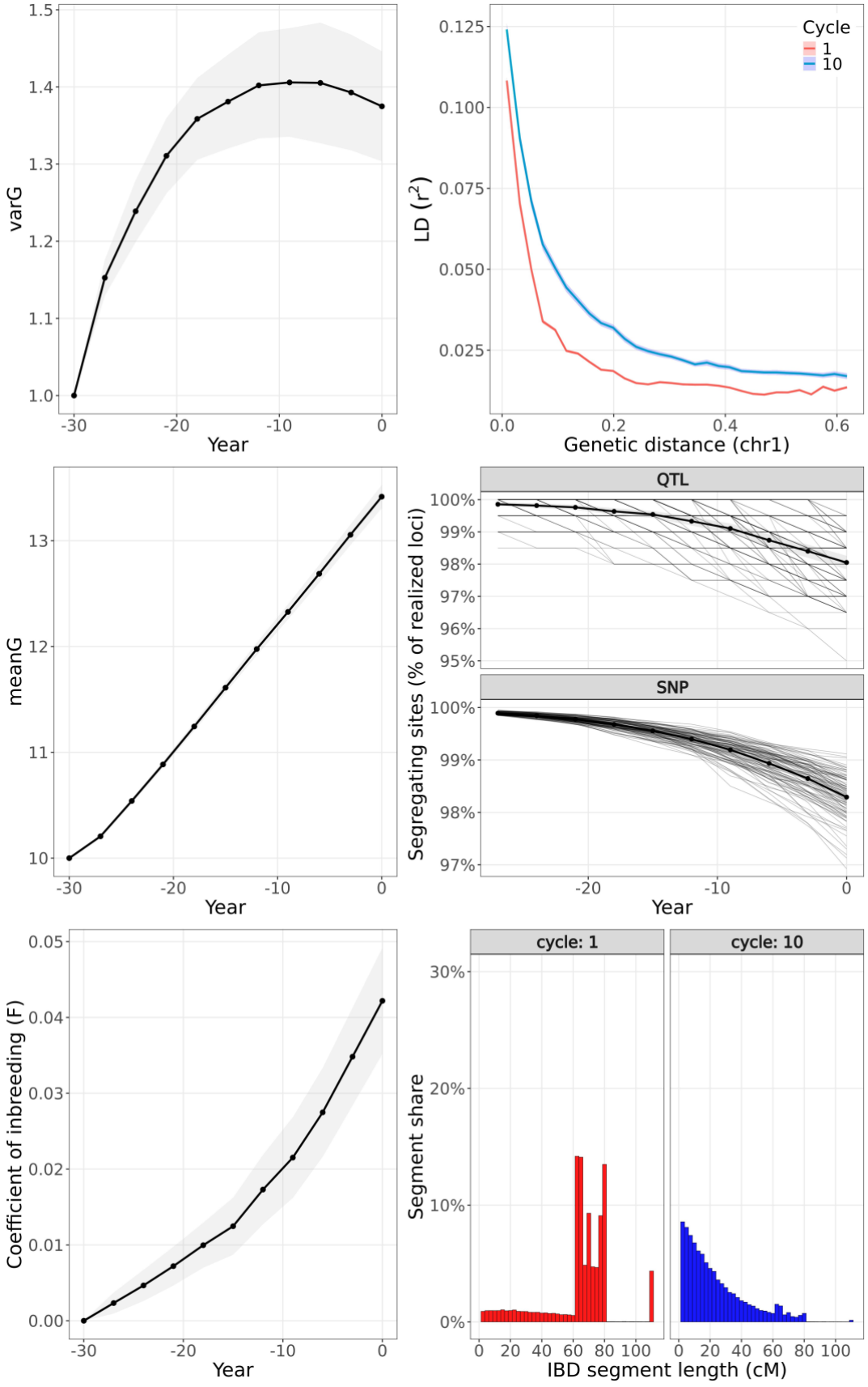

varG
Year
LD (r^2)
Genetic distance (chr1)
Cycle
1
10
meanG
Year
QTL
SNP
Segregating sites (% of realized loci)
Year
Coefficient of inbreeding (F)
Year
cycle: 1
cycle: 10
Segment share
IBD segment length (cM)

Figure 2. Evolution of genetic and genomic diversity indicators during the apple burn-in phase. The burn-in consisted of ten preliminary phenotypic-selection cycles used to initialize the parental population before comparing breeding strategies. Years are expressed relative to the end of the burn-in, with year 0 corresponding to the parental population used to start the scenario simulations. The panels show the evolution of additive genetic variance (varG), mean genetic value (meanG), and the parental-population inbreeding coefficient (F); the decay of linkage disequilibrium (LD, r^2) with genetic distance on chromosome 1 in cycles 1 and 10; the proportion of initially realized QTL and SNP loci that remained segregating over time, with thin lines representing individual simulation replicates and thick lines representing the mean; and the distribution of identical-by-descent (IBD) segment lengths in cycles 1 and 10. Across the burn-in, directional selection increased the population mean and inbreeding, while genetic variance initially increased and subsequently stabilized. The proportion of segregating loci declined moderately, particularly for QTL, and the IBD distribution shifted toward a larger proportion of short segments by cycle 10. LD decayed rapidly with genetic distance in both cycles, although LD was generally higher in cycle 10. Shaded ribbons represent variation among simulation replicates.

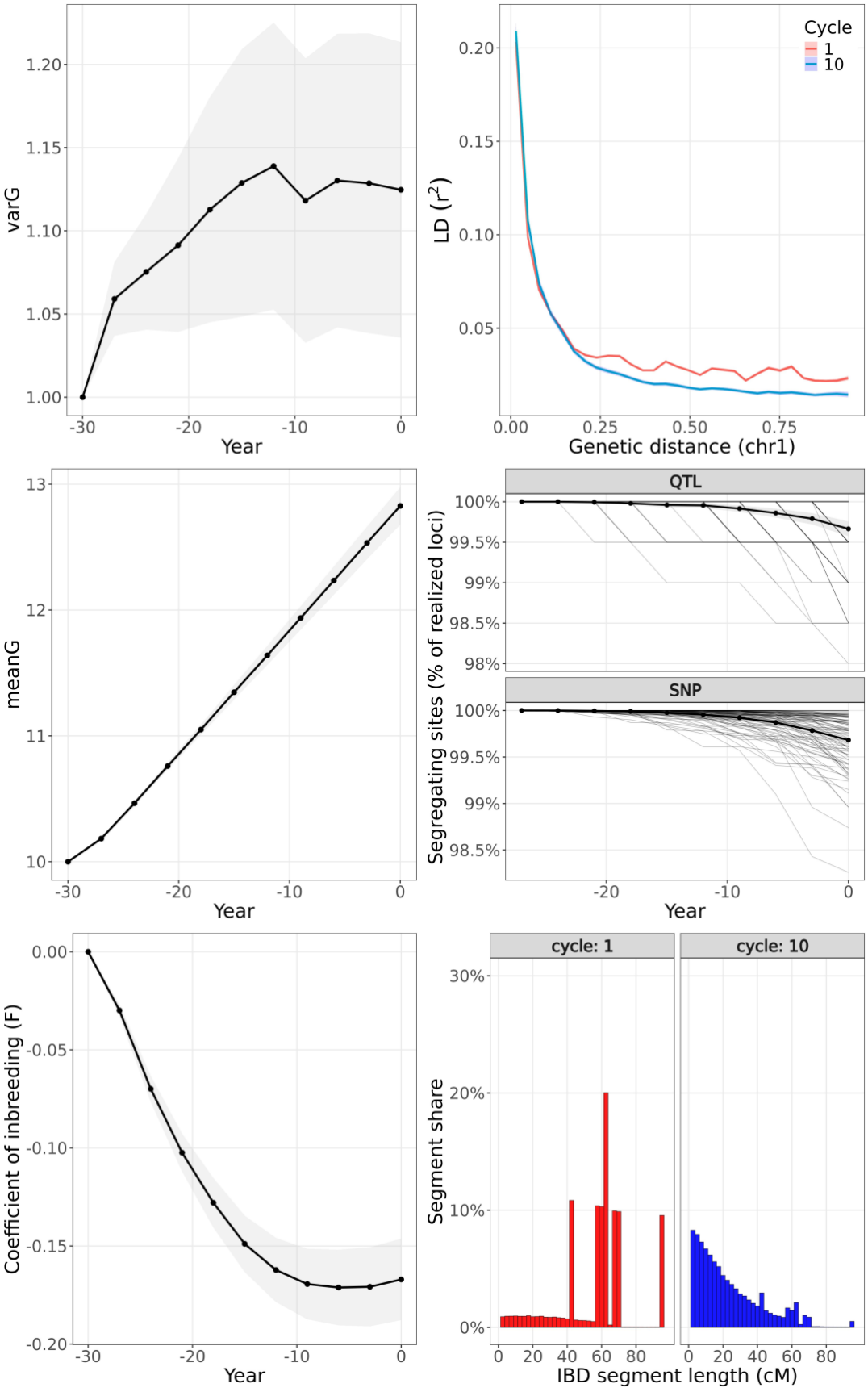

varG
Year
LD (r^2)
Genetic distance (chr1)
Cycle
1
10
meanG
Year
QTL
SNP
Segregating sites (% of realized loci)
Year
Coefficient of inbreeding (F)
Year
cycle: 1
cycle: 10
Segment share
IBD segment length (cM)

Figure 3. Evolution of genetic and genomic diversity indicators during the peach burn-in phase. Across the burn-in, directional selection increased the mean genetic value, while genetic variance increased during the first cycles and then remained broadly stable. The proportion of segregating loci declined only slightly, indicating limited loss of QTL and SNP polymorphism during initialization. LD decayed rapidly with genetic distance in both cycles and was generally lower in cycle 10 at intermediate and long distances. The IBD distribution also shifted toward a higher frequency of short segments by cycle 10. The negative values observed for the estimated inbreeding coefficient indicate that this statistic was centered relative to the founder population and should therefore be interpreted as outbreeding. Shaded ribbons represent variation among simulation replicates.

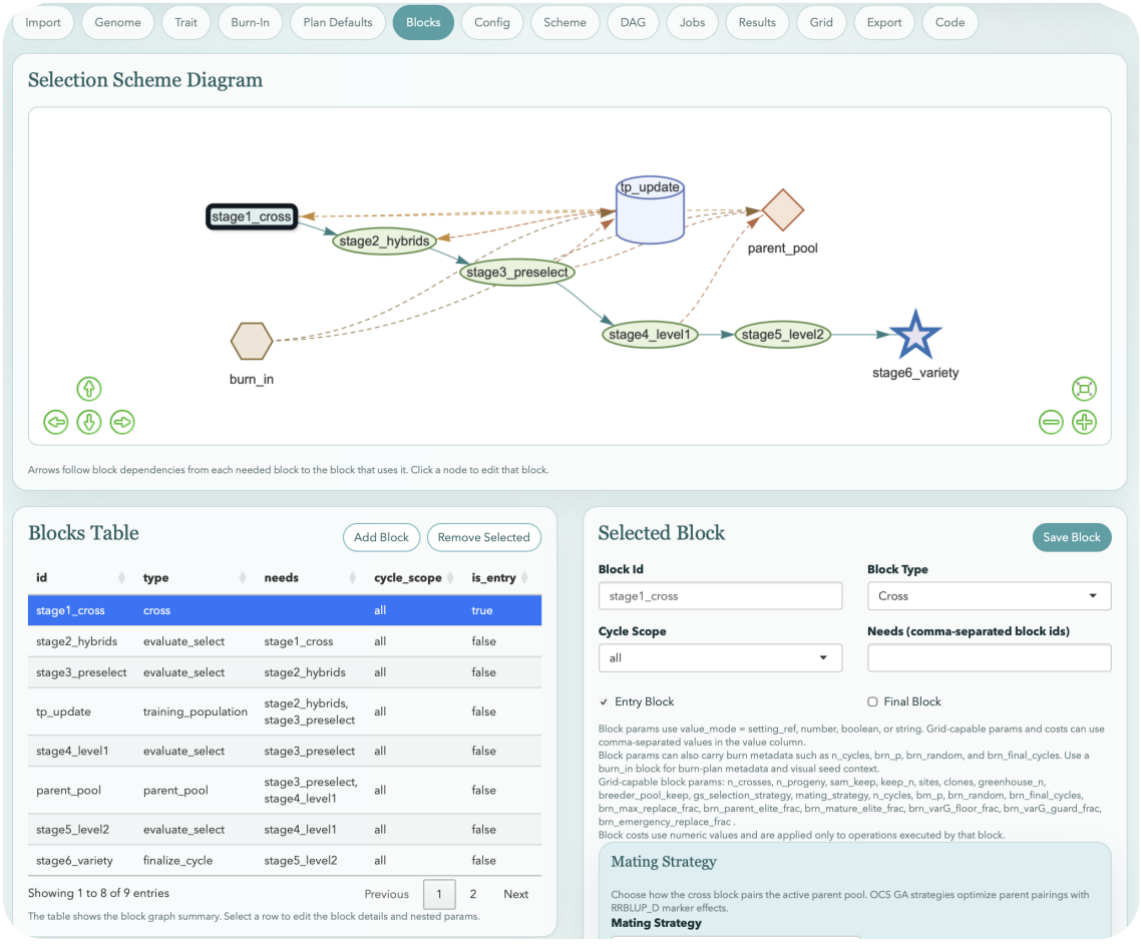


Figure 4. Shiny UI developed for this project. The Shiny App was especially valuable during the exploratory phase to test configurations and schemes and served as the foundation for these scripts.