BREEDING SCHEME DEVELOPMENT AND OPTIMIZATION DURING NEO-DOMESTICATION AND WIDE HYBRIDIZATION

A DISSERTATION SUBMITTED IN FULFILLMENT OF A REQUIREMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

TROPICAL PLANT AND SOIL SCIENCES

UNIVERSITY OF HAWAIʻI AT MĀNOA

December 2023

By

Nathan John Fumia

Dissertation Committee:

Michael Kantar, Chairperson
Tai Maaz
Rosana Zenil-Ferguson
Marnin Wolfe, Ad-Hoc
Daniel Rubinoff

Contents

List of Tables

List of Figures

List of Abbreviations

CES — Crossing, Evaluation, Selection
koa — *Acacia koa*
cacao — *Theobroma cacao*
stevia — *Stevia rebaudiana*
FOXY — *Fusarium oxysporum f. sp. Koae*
RCBD — Randomized Complete Block Design
PDA — Potato Dextrose Agar
PDB — Potato Dextrose Broth
QTL — Quantitative Trait Loci
LMM — Linear Mixed-Model
SGF — Sweet Green Fields, Inc.
HARC — Hawaii Agriculture Research Center
IM — Intermated Population
BLUP — Best Linear Unbiased Prediction
BLUE — Best Linear Unbiased Estimation
GS — Genomic Selection
OCS — Optimum Contribution Selection
Pedigree-OCS — Pedigree Based OCS
GS-OCS — Marker Based OCS
GBLUP — Genomic BLUP
GEBV — Genomic Estimated Breeding Value
TOI — Trait of Interest
dsTOI — Domestication Syndrome TOI
ssTOI — Stevia Specific TOI
RebA — Rebaudioside A
ROI — Return on Investment
Ne — Effective Population Size
z — trait/character
PRS — Phenotypic Recurrent Selection
MxAv — Maximum Avoidance Selection
h^2 — Narrow-Sense Heritability
si — Selection Index

Overview

Global food systems are under increasing pressure. There is an increasingly stochastic climate with future projections showing declines in major food crop production across the major growing locales. In the future, the geographic regions that are projected to be suitable for staple production are expected to shift, which may lead to the need to abandon current production regions and/or shift to new species for cultivation. Many of the potential new species to replace current systems or fill novel niches include ∼30,000 edible plants worldwide; however, currently ∼150 are cultivated at large scale across the world. These domestic and semi-domestic plant species span 160 taxonomic families with a total of ∼2,500 species having undergone some extent of domestication. Domestication is a process by which a wild organism shifts to a form more adapted for human use, typically through the acquisition and subsequent fixation of traits, commonly termed the domestication syndrome. Neo-domestication is the attempt to re-domesticate or newly domesticate wild and semi-wild species leveraging modern breeding techniques in a strategic framework. These programs use large phenotypic and/or genotypic data sets to efficiently select breeding parents to maximize progeny gain of typically quantitative traits. Neo-domestication foci include fixation of simple traits, via variance reduction, that typically hinder cultivation and breeding for complex traits that improve marketability. The increasing pressure on food systems and breadth of options creates a situation of strategic uncertainty, specifically two questions:

1) Which species do we choose?

2) How do we increase the adaptability of said species to the agroecosystem?

This dissertation aims to systematically answer the second question to understand the impact of breeding schemes on the pace of adaptation. Specifically breeding scheme development encompasses the parametrization in breeding cycle components (Crossing, Evaluation, Selection) by leveraging empirical data to train stochastic models. First, empirical data are integrated with simulation output to test specific use-cases, this is followed by optimization of genetic gain. The empirical case study chapters (1-3) rely upon the concepts of population improvement through artificial selection: an iterative process of generational selection to increase in favorable alleles in the population (Figure 1). The goal is to increase the probability of extracting a superior cultivar from the population. The increase in favorable alleles drives genetic gain, which is a product of additive genetic variation within the population, selection intensity, and selection accuracy (Figure 1). Optimizing breeding cycle components has beneficial effects on the genetic gain components. However, this framework has not been applied to wild and semi-domesticated breeding towards domestication. The last chapter (4) integrates the effects of these case studies to create a framework for understanding the expected rate of gain for potential species with known biological characteristics.

In Chapter 1, the first breeding cycle component addressed is crossing. Crossing includes the following parameters: number of parents, number of crosses, number of progeny, type of

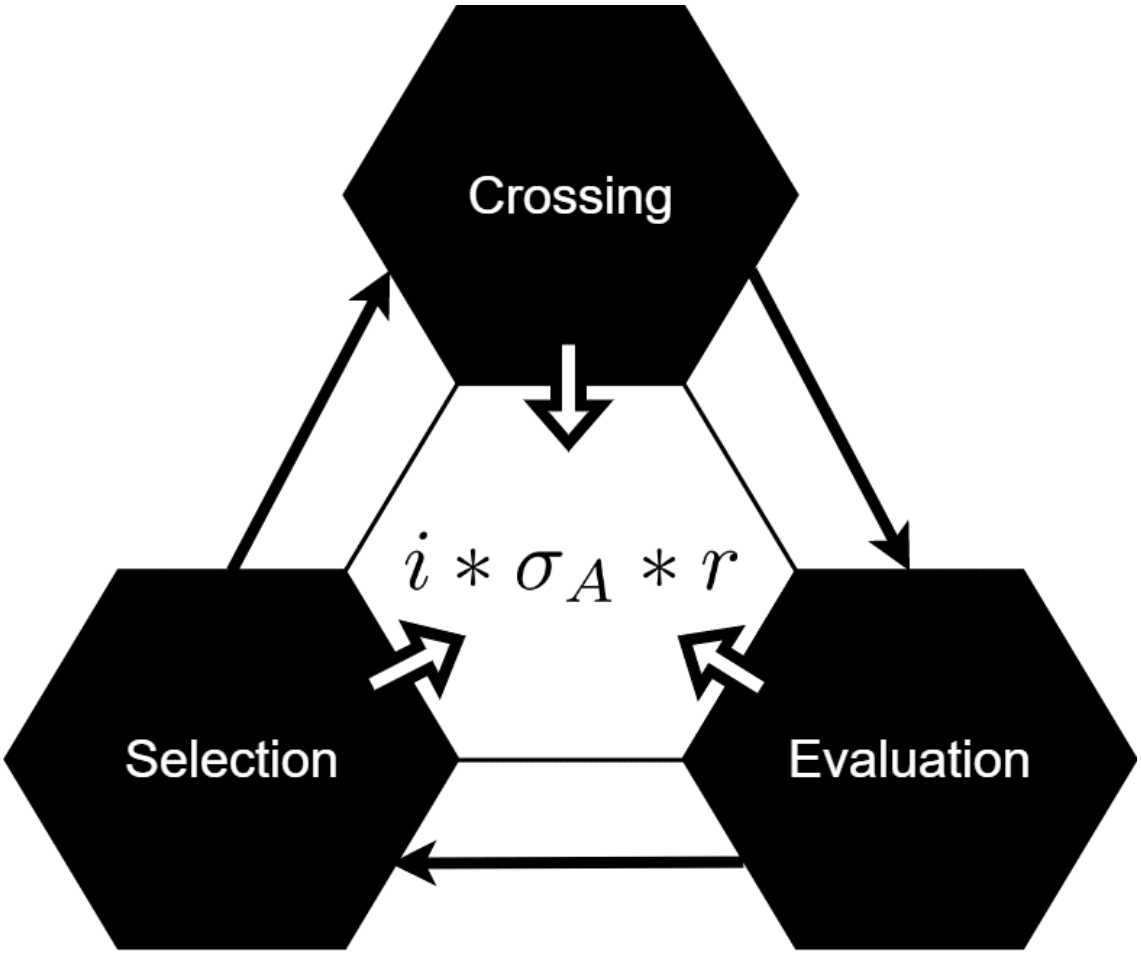
cross, and mate allocation. To parse parametric effects in crossing on gain, the wild-endemic tree species *Acacia koa* (koa) was used. Koa is an excellent system for understanding crossing parameters of number of parents and number of progeny effects on gain in seedling vigor with disease resistance constraint. The koa system provides a clear set of experiments to validate the influence of breeding population size and number of progeny on phenotypic gain in simulation results of seedling vigor when constrained by varying levels of disease.

The second chapter focuses on the next component of the breeding cycle, evaluation. This component includes the following parameters: number of locations, levels of replication, number of checks, experimental design, and subsampling. To parse parametric effects in evaluation on gain, the tropical tree species *Theobroma cacao* (cacao) is used. Cacao production spans the globe, but production varieties vary from developed (hybrids) to semi-domesticated (open-pollinated landrace). The cacao system provides a clear set of experiments to understand the precision necessary to observe significant differences between varieties within different domestication status groups. The investigation into the deviation of phenotypic measurements from the most precise (full sub-sampling of plot) to least precise (single sample of plot) will facilitate an understanding of the appropriate level of precision needed when using different types of germplasm in a neo-domestication breeding program.

The last aspect of the breeding cycle, selection, is addressed in chapter 3. This component includes the following parameters: percentage selected (intensity), selection method (culling, index), selection unit (families, lines, parents), and selection criteria (phenotypic, genotypic, breeding values, index). To parse parametric effects in selection on gain, the subtropical herbaceous shrub species *Stevia rebaudiana* (stevia) is used. Stevia production spans the globe, but production varieties are considered non-adaptable to the agroecosystem (semi-domesticate) and lack classic domestication syndrome traits, including reduced dormancy, branching, and photoperiod sensitivity. The stevia system provides a clear set of experiments to understand the influence of phenotypic recurrent selection on domestication traits across multiple generations.

Although these systems differ by life-history and domesticated status, they provide key insight into the components of the breeding cycle with varying trait complexities. The simulation varying crossing parameters can take the empirical estimates from Koa and expand intuition around other parameters during crossing (e.g., varying number of crosses). The simulation varying evaluation parameters can take the empirical estimates from Cacao and expand intuition around other parameters during evaluation (e.g., level of replication or environmental evaluation). The simulation varying selection parameters can take the empirical estimates from Stevia and expand intuition around other parameters during selection (e.g., alternative selection criteria). Each variable is therefore grounded in crucial empirical estimates derived from experiments in each case-study crop. The input for these case studies provides the knowledge and intuition for the final chapter, which synthesizes these expectations into species agnostic neo-domestication breeding schemes. This synthesis uses replicated and varying stochastic simulations across the breeding cycle components to provide an estimate of potential gain for any domestication scenario. These new breeding schemes are returned to reality by applying known cost structures to key parameter decisions (e.g., cost of phenotyping versus genotyping; cost of subsampling versus plot level analysis). Mixed-model analysis is then applied to estimate breeding cycle component parameter values which maximize the return on investment for given traits and targets.

Figure 1: Schematic of how the components of the breeding cycle each play a role in the response to selection, or genetic gain.



Chapter 1: *Acacia koa* seedling disease tolerance driven by breeding orchard size - informing breeding cycle crossing expectations

1.1 Introduction

Acacia koa Gray (hereafter, koa) is an endemic timber tree species of the Hawaiian archipelago with important cultural and ecological relevance (Baker, 2009). Historically used for ocean-voyaging canoes and is presently used as a high value input in fine furniture, jewelry, and musical instruments. Past deforestation for agriculture and ranching during the 19th and 20th centuries, along with its exquisite curly grain and rich coloration has driven the price to $125 per board foot and estimated annual gross value of $20-30 million (Yanagida et al., 2004). Further, koa is a critical canopy tree providing habitat for endangered native birds and epiphytic plants (Pejchar et al., 2005, Baker, 2009).

Koa is an obligate outcrossing (cross-pollinating), autotetraploid species (2n=52) (Carr, 1978, Shi, 2003). The species arose from a genome duplication event of the Australian species *Acacia melanoxylon* (Brown et al., 2012, Le Roux et al., 2014). Colonization of the Hawaiian archipelago by koa likely occurred millions of years ago based on pollen records (Hotchkiss and Juvik, 1999) and host-specific endemic insect species (Gagne, 1979). Phenotypic and genotypic variation is found between and within different populations through the Hawaiian Islands (Wagner et al., 1990, Sun et al., 1996, Adamski et al., 2012, Dudley et al., 2020). Variation can also be found by eco-region (primarily wet windward or dry leeward populations) and elevation, yet these sub-groups still hierarchically cluster by island (Dudley et al., 2017).

Compounding the impact of deforestation on population decline is the rise of *Fusarium oxysporum f. sp. Koae* (FOXY), commonly known as koa wilt disease (Gardner et al., 1980, Anderson et al., 2002, Dudley et al., 2007), decimating the remaining forest by clogging xylem flow of infected, susceptible trees. However, varying levels of resistance to the disease-causing pathogen can be found in each population of koa (Dudley et al., 2015) prompting investigation into disease resistant seed production for koa orchard and reforestation efforts (Dudley et al., 2020). The cultural, ecological, and economic importance of koa is the driving force behind facilitating range expansion of the species by increasing tolerance to the disease in distinct koa populations, providing a mechanism by which koa realizes a domesticated state capable of cultivation in the agroecosystems and restoration of the ecosystems of Hawai'i. The traits essential for domestication vary depending on the organism, its life-history and likely agroecosystem. Koa, as a long-lived perennial with agricultural and ecological significance, requires traits that facilitate range expansion, one of which is the ability to tolerate or resist infection by FOXY. Although disease resistance typically has simple genetic architecture

(Jungers et al., 2023), tolerance is often exemplified through improvements in different more complex traits, such as vigor.

Population improvement through artificial selection requires optimization of genetic gain for different traits and tailored towards the end goal (Gaynor et al., 2017). This is achieved through an iterative process of generational increase in favorable alleles in the population under selection, acting to increase the probability of extracting a superior cultivar from the population (Cobb et al., 2019, Van Tassel et al., 2020). The increase in favorable alleles drives genetic gain, which is a product of additive genetic variation within the population, selection intensity, and selection accuracy. Each gain component can be increased through different methods and technology applied across the breeding cycle (Hickey et al., 2017, Wallace et al., 2018, Wartha and Lorenz, 2021). Population improvement relies upon adequate parametrization in the breeding cycle of crossing, evaluation, and selection (Covarrubias-Pazaran et al., 2022).

The focus of this study is the effect of crossing on a wild species being selected for a domesticated form, including decisions such as number of parents, number of crosses, and number of progeny to continue. Moreover, this chapter accounts for the interplay of traits, seedling vigor and disease resistance – a common occurrence during neo-domestication. Here general knowledge on how to begin breeding scheme development for wild species (neo-domestication) is developed through the use of koa as a novel system to understand crossing and breeding population size during incipient domestication. The alteration of breeding population size in the koa orchard is through thinning of individuals, a form of pollen control in the population. These individuals are thinned for varying reasons, one of which being a low durability of resistance which causes trees to succumb to disease over time. As the breeding population size decreases, and therefore diminished background genetic variability, the progeny survival probability is increased by removal of susceptible types from the breeding population. This is an opportunity for insight to develop expectations of population improvement through augmented crossing parameters to inform situational changes through breeding cycles in neo-domestication programs (Cobb et al., 2019, Covarrubias-Pazaran et al., 2022).

1.2 Materials and Methods

1.2.1 Germplasm and breeding

Individual wild accessions were sourced from all major islands in the Hawaiian archipelago (Hawai'i, Maui, O'ahu, Kaua'i) at dry-leeward and wet-windward locations in the late 1990s where seed was grouped into half-sibling bulk seed (Supplemental Figure 1). Disease screening following methods reported in (Dudley et al., 2015) were completed for each half-sibling bulk where maternal lineages were removed for low progeny disease resistance (proportion survival <60%). Resistant maternal half-sibling seedlings (proportion survival >60%) were sourced from the wet-windward O'ahu population to include in the randomized complete block design orchard in 1996. It was not until 12 years later that seed was collected, following the first orchard thinning. Thinning in the orchard is done to improve tree structure, increase diameter at breast height (DBH), and remove individuals with low durability of resistance and low or

no seed production. This process has been repeated three more times for a second, third, and fourth thinning with subsequent seed collections (Supplemental Figure 1).

1.2.2 Trials

To quantify the effect of breeding population size on seedling disease resistance, bulk seed from each thinning group (1^{st}, 2^{nd}, 3^{rd}, 4^{th}) were placed into a randomized complete block design in greenhouse for disease screening compared against wild seed (native range – Ko'olau Mountains). Disease screening methods follow those used in 1996 and (Dudley et al., 2007). FOXY inoculum, derived from wild pathogenicity trials (Dudley et al., 2007) is used as a dried fine powder and mixed thoroughly with growing media. The inoculum was a cocktail of 9 highly virulent races of FOXY, identified and isolated by Dr. Susan Schencke and Nick Dudley to estimate broad tolerance to the variety of virulent races likely to be encountered in the agroecosystem (Dudley et al., 2007). Once koa seed has germinated (stratified and soaked), the germinated seed is placed into inoculated media and watered, held growing until visual signs of wilt and/or death is observed. Death by FOXY is quantified by including a split-plot to our RCBD where half of the seedlings were planted into inoculated media and half into sterile media. Therefore, proportion of group survival will be used as the measure of disease resistance per population.

Koa seedlings from each seed group and split were also sampled for pathogen confirmation. Each of the 33 sampled seedlings had its root collar, which included a small portion of the root and stem, severed and surface sterilized with 10% Clorox bleach solution and rinsed twice in sterile water. These segments were transferred to $\frac{1}{4}$ potato dextrose agar (PDA) petri plates and incubated at 25C for 3 days. Once mycelial growth was observed from each sample, hyphal tips were taken for each seedling that appeared to be Fusarium. These cultures were grown for a week to get sufficient growth of mycelium for storage. Representative cultures were grown in potato dextrose broth (PDB) for 7 days for extraction of total DNA using a Zymo plant/fungal DNA extraction kit. Polymerase chain reaction (PCR) was used to amplify the rpb2 locus for these isolates and sent for sanger sequencing in both the forward and reverse directions at Eurofins genomics. Sequences were trimmed and aligned using Geneious prime and consensus sequences were BLASTed to NCBI GenBank database for identification.

1.2.3 Analysis

Data collected during this trial include weekly survey of death (timing of death) and seedling height at trial end (120 days). To investigate the effect of breeding population size on progeny seedling disease resistance, we employ binomial regression with a logit-link function through a generalized mixed-effects model using the R Statistical Software package 'lme4' (Bates et al., 2014) to make predictions about the probability of survival of a given seed dependent on its source group. The formula uses the ratio of planted seeds to total death as the response variable with random variables of block and fixed variables seed group and split treatment. Next, we use a two-sample t-test to compare means of sterile and fusarium treated (split treatment) because disease pressure degrades plant vigor, especially affecting seedling stage (R Core Team, 2021). Moving past the t-test, we apply a linear mixed-model approach to estimate seed group and treatment effects on seedling vigor at 120 days (height in cm) while

removing random blocking effects.

1.2.4 Simulation integration and comparison

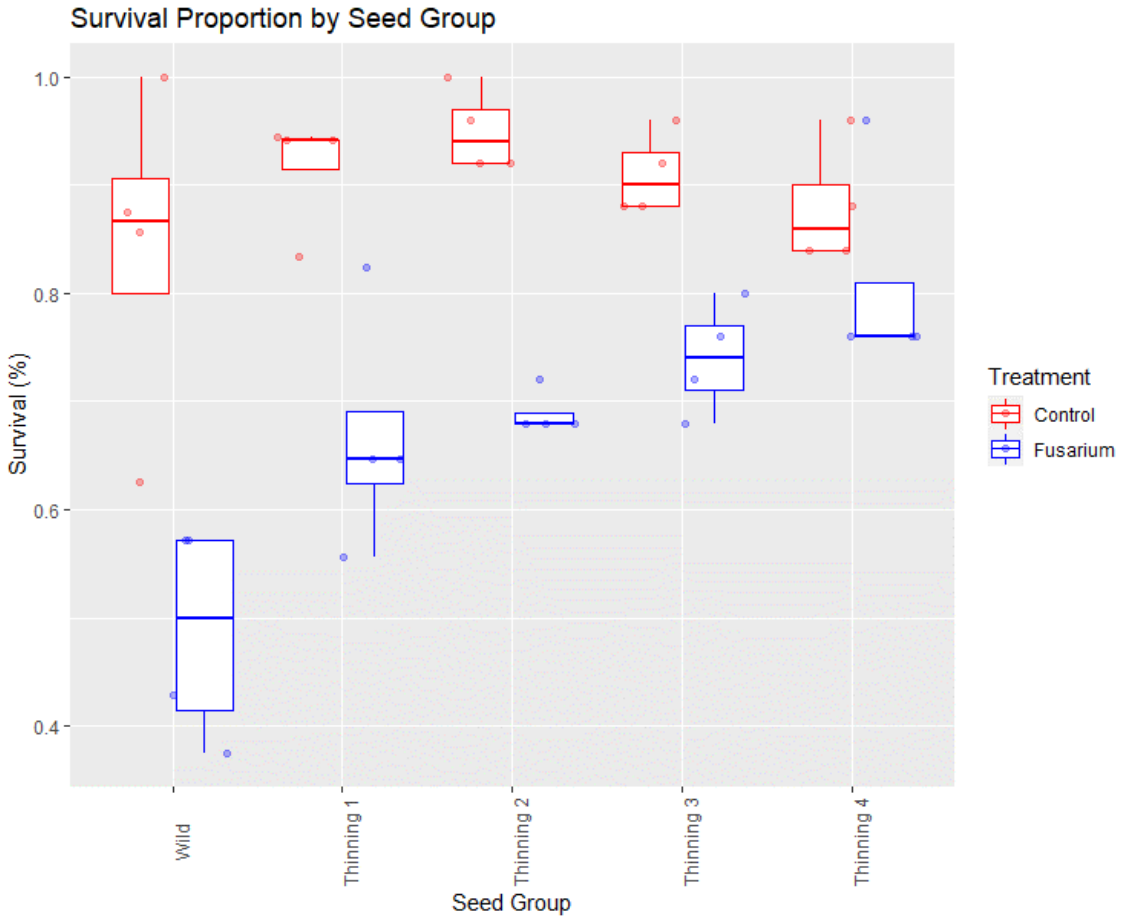
Once mixed-model analysis estimates seed group effect and predicts residual errors, we move to integrate to stochastic simulation using the R Statistical Software package 'AlphaSimR' (Gaynor et al., 2021) for future projection and estimating genetic complexities. More specifically, we adjust the founding population parameters to match estimates of our different seed sources to detangle population size and resistance influence on seedling vigor. For future projection, we simulate alternative genetic complexities for seedling vigor as simple oligogenic (8 QTL), complex oligogenic (20 QTL), and polygenic (100 QTL) over 10 cycles of selection. Designing the founders takes species specific information such as 13 chromosomes in koa, a genome duplication to simulate autotetraploidy, as well as specifying mean and variance for the trait along each seed group (Wild through GroupE, Table 2). Once trait is specified under a given genetic architecture, phenotypes are estimated by taking into account the additive genetic architecture with the residual error, estimated through mixed-model (Table 2). We then specify factors relevant to the breeding cycle, such as number of parents (# of maternal parents for each group, Table 2), number of crosses per parent (parents-1), and number of progeny from each parent (varied according to number of parents to have ∼500 total progeny). Number of parents follows empirical seed group amounts (Table 2) while number of progeny per parent is altered to maintain ∼500 total progeny each cycle (Wild: 53x10; First: 43x12; Second: 18X28, Third: 17x30; Fourth: 16x32). The founding population (cycle 0) therefore matches wild seed group linear mixed-model estimates while each seed groups linear mixed-model estimates form cycle 1 and serve as the backbone for 10 cycles of selection within each group. The 10 cycles of selection are selected using truncation selection for number of parents with the highest phenotypic observation. But, important to this simulation is the presence of disease, which we simulate through random selection and removal of breeding candidates for death at the proportion of death (1-survival probability). In effect, each generation there are randomly selected candidates removed from the population to integrate disease susceptibility and death, where the proportion of selected is calculated through the survival probability estimated in each seed group. The simulation is replicated 100 times and the mean (per cycle) phenotypic value and variance, along with their respective standard errors, are output.

We move to investigate the effects of breeding population size, number of progeny, and survival probability on seedling vigor gain through 10 cycles of selection. To do so, we select polygenic (100 QTL) to serve as the genetic architecture of seedling vigor, with founding population and cycle 1 being formed the same way as previously but altering number of parents, progeny, and survival probability. Therefore, we set 3 comparative schemes: (1) seed group survival probabilities remain the same as empirical while setting constant - number of parents (43) and progeny (12); (2) seed group survival probabilities remain the same as empirical while setting constant – number of parents (16) and progeny (32); and (3) seed group survival probabilities are inverted (First to Fourth, Second to Third, vice versa) and seed group number of parents and progeny remain the same as empirical. These alterations will inform our understanding of the effect of survival probability on vigor gain through 10 cycles of selection and the effect of breeding population size to mitigate these effects.

1.3 Results

The size and status of a breeding population has direct influence on the frequency of important traits like disease tolerance, vigor, and survivability in progeny seedlings. Differences between the sterile and disease treatments were observed during trial execution in survival and vigor. The proportion of survival in sterile groups ranged from 85-95%, but with the application of FOXY to media during planting there is variable effect by thinning group (Figure 2). Wild source seed has the lowest tolerance to disease (mean: 50%) while the seed collected following the 4th thinning of the seed orchard has the highest (mean: 75%). Each seed group between the wild and the 4th thinning exhibit incrementally improved survival (Figure 2).

Figure 2: Proportion of seedling survival by seed source group. Comparison of sterile split (no FOXY) and fusarium (FOXY) tolerance.



1.3.1 Case Study of Koa Domestication

1.3.1.1 Generalized mixed-model estimates of survival probability

Generalized binomial mixed-effects model identifies seed from the second, third, and fourth thinning as being significantly higher proportion survival as compared to the intercept (wild seed; $p<0.05$). However, estimates of these groups are not different from one another, improving log-odds by 0.66 (0.32), 0.67 (0.32), and 0.77 (0.33) over the wild seed intercept of 1.79 (0.30), respectively (Table 1). Furthermore, when we apply the logistic link function to the log-odds, sterile probability of survival is 85.7% in wild seed with the largest increase in survival of the 5-year-old orchard (1st thinning, Group B) to 90.8% (Table 1). Probability of survival is further increased in the 6-year-old orchard (2nd thinning, Group C) to 92.0% with minimal increase in following years and orchard thinnings to the 9-year-old orchard (4th thinning, Group E) to 92.8%. Application of fusarium (treatment) reduces predicted log-odds of survival by 1.13 (0.20; $p<0.0001$). Logistic link function finds treatment effect of fusarium disease probability of survival dropping in every group, with the largest decrease in survival estimated in wild seed by 20% to 65.9% survival and the smallest decrease in survival estimated in the 9-year-old orchard (4th thinning, Group E) by 12% to 80.7% survival (Table 1).

Table 1: Generalized linear mixed-model log-odds estimates and probability of survival by seed group under sterile (untreated) and fusarium (treated) conditions.

GLMM Coefficients	**Seed Group** (Maternal #)	**LogOdds** Estimate (SE)	**Probability** Control	**LogOdds Fusarium** Estimate (SE)	**Probability** Fusarium
Intercept	Wild Seed 2012 (53)	1.790 (0.301)	0.857	-1.130 (0.203)	0.659
B Effect	1st Thinning (43)	0.498 (0.338)	0.908		0.761
C Effect	2nd Thinning (18)	0.659 (0.323)	0.920		0.789
D Effect	3rd Thinning (17)	0.668 (0.325)	0.921		0.790
E Effect	4th Thinning (16)	0.771 (0.330)	0.928		0.807

1.3.1.2 Linear mixed-model estimates of seedling vigor

Survival proportion is not the only effect of FOXY on koa seedlings where vigor (seedling height) of remaining seedlings (end of trial) is diminished across all seed source groups (Figure 3). Welch's two sample t-test identifies significant ($p<0.0001$) differences between the split – sterile versus FOXY applied, with mean height of sterile 21.53 cm and of FOXY treated 17.70 cm. Linear mixed-model is used to estimate vigor (seedling height) at 120 days post germination in sterile (untreated) and fusarium (treated) conditions (Table 2). The unexpected trend identified in data exploratory visualization (Figure 3) is corroborated using mixed-model, where the surviving wild seed has the highest vigor estimate (19.28 cm in height) out of all groups with Group E (4th thinning) as the next highest estimate (18.69 cm; Table 2). It can be inferred that this trend is exhibitive of disease tolerance in orchard groups where rather than fusarium causing death in seedlings (lack of resistance observed in wild seed; Table 1), seed derived from orchard groups have diminished growth and vigor (Table 2).

1.3.2 Stochastic simulation for projected gain under variable crossing parameters

Figure 3: Seedling vigor (height) by seed source group. Comparison of sterile split (no FOXY) and fusarium (FOXY) tolerance.

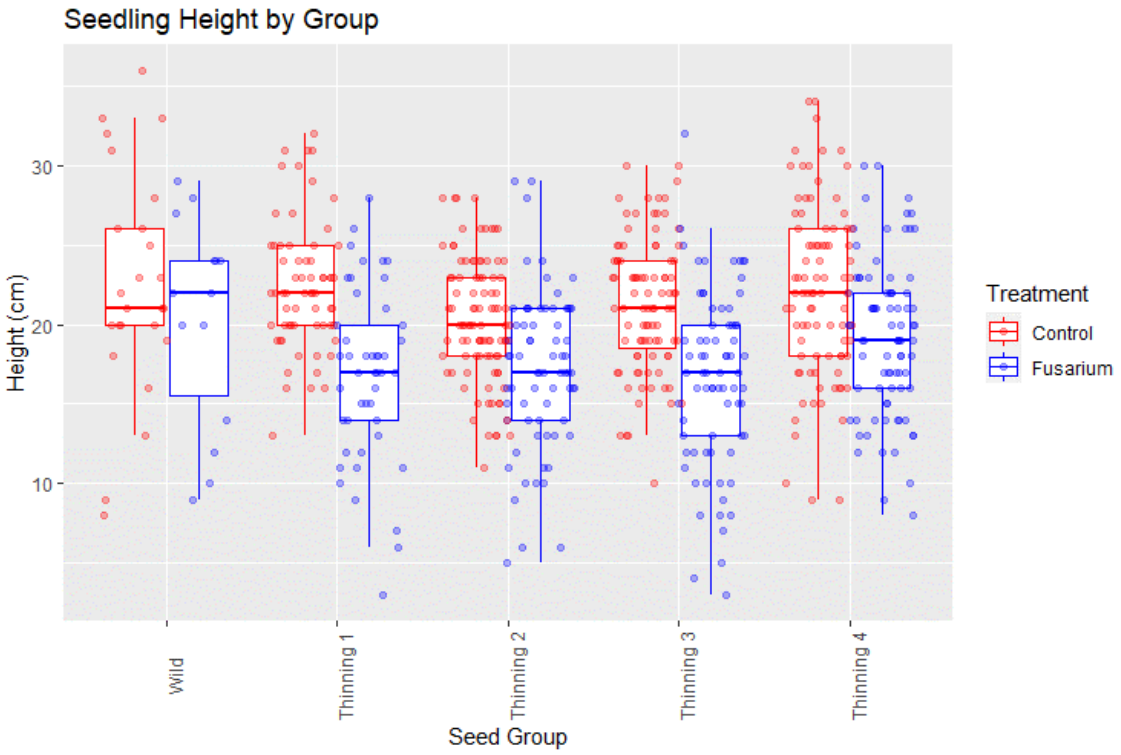


Table 2: Linear mixed-model estimates of seedling vigor, quantified as height (cm) at 120 days post-germination, under sterile (untreated) and fusarium (treated) conditions.

LMM Coefficients	**Seed Group** (Maternal #)	**Vigor Control** Estimate (SE)	**Vigor Fusarium** Estimate (SE)	**Vigor Fusarium** Estimate
Intercept	Wild Seed 2012 (53)	23.112 (0.850)		19.279
B Effect	1st Thinning (43)	-1.371 (0.923)		17.908
C Effect	2nd Thinning (18)	-2.516 (0.882)	-3.833 (0.392)	16.763
D Effect	3rd Thinning (17)	-2.215 (0.882)		17.064
E Effect	4th Thinning (16)	-0.592 (0.880)		18.687

Stochastic simulation was used to project gain in seedling vigor with variable seedling survival via FOXY resistance. We begin with parameters matching empirical for each of our four orchard seed groups to project gain in vigor through 10 cycles of selection with varying genetic architecture complexity. Simple oligogenic control (8 QTL) of seedling vigor finds a rank change in the top two vigor estimate seed groups (4th thinning: 18.69 & 1st thinning: 17.91) after 10 cycles of selection (Figure 4A). The 1st thinning seed group gains phenotypically 41% over 10 cycles of selection, compared to 29% in 4th thinning seed group. This is a likely result of more genetic variation within the 1st thinning seed group (43 breeding parents). These groups have a projected split at generation 6, where standing variation in each population has diminished by 84% in the 4th thinning group versus 58% in the 1st thinning group (Figure 4B). Increase in genetic complexity to complex oligogenic (20 QTL) of the simulated trait finds no overlaps in genetic gain in 10 cycles of selection with seed groups ranked in accordance with estimates used (Figure 4C). It appears 1st thinning seed group will overtake the 4th thinning seed group in the 11th cycle, especially considering the loss of genetic variation is almost 10% greater in the 4th thinning group, meaning more genetic variation for selection is present (Figure 4D). Another increase in genetic complexity to polygenic (100 QTL) of the simulated trait finds overlaps in genetic gain in 10 cycle of selection with seed groups ranked according to their survival probabilities, with the worst in gain being the first (67%) and the best in gain being the fourth seed group (81%), highlighting the potential for disease resistance to maintain genetic variation becomes more important to gain in primary traits as the architecture of those traits becomes increasingly complex (Figure 4E). The effect of a rather simple trait, disease resistance, on the gain in a complex trait, vigor (simulated: 100 QTL), raises some questions regarding whether this effect may be mitigated through breeding population size.

We therefore implement 3 different schemes by altering either seed group crossing parameters (number of parents and progeny) and/or survival probability. The first scheme was designed to test the effect of survival probability on projected seedling vigor gain when using the different seed groups (differ by cycle 1 estimates: Table 2) while maintaining a larger breeding population size (43 parents). Here we observe increases in final gain comparative to the seed groups survival probability: First (67%) gains 7.06 cm in height, Second (69%) gains 8.14 cm, Third (74%) gains 8.73 cm, and Fourth (81%) gains 8.92 cm (Figure 4A). Moreover, this change in breeding population size from 16 parents to 43 parents increases variance in the 4th thinning seed group by almost 10% in the 10th cycle of selection (Figure 4B). The second scheme was designed to test the effect of survival probability on projected gain when using a smaller breeding population size (16 parents) for each seed group.

Results of gain are more variable in the second scheme, but seed groups finish 10 cycles of selection in the same order as their cycle 1 estimates (Table 2 & Figure 4C). The more noticeable effect of the decrease in breeding population size is the loss of variance, where in the 10th cycle for each seed group has more variation in the second than in the first scheme: First loses 67% versus 75%, Second loses 66% versus 77%, Third loses 61% versus 76%, and Fourth loses 66% versus 73% (Figure 4B/D). The third scheme was designed to test whether the effect of survival has a proportional effect to the variable breeding population sizes used empirically. Despite differing starting points (seed group vigor estimates; Table 2), when survival probabilities are inverted (decreasing) and breeding population parameters are kept the same as empirical (decreasing), final projected gain of height in cm across all

four groups is almost within 1 unit (First: 26.92 cm & Third: 25.73 cm; Figure 4E). More variation (30%) is present in the first seed group breeding population than the alternatives, highlighting the importance of a large breeding population size (43 parents) and high degree of disease resistance (81%) (Figure 4F).

1.4 Discussion

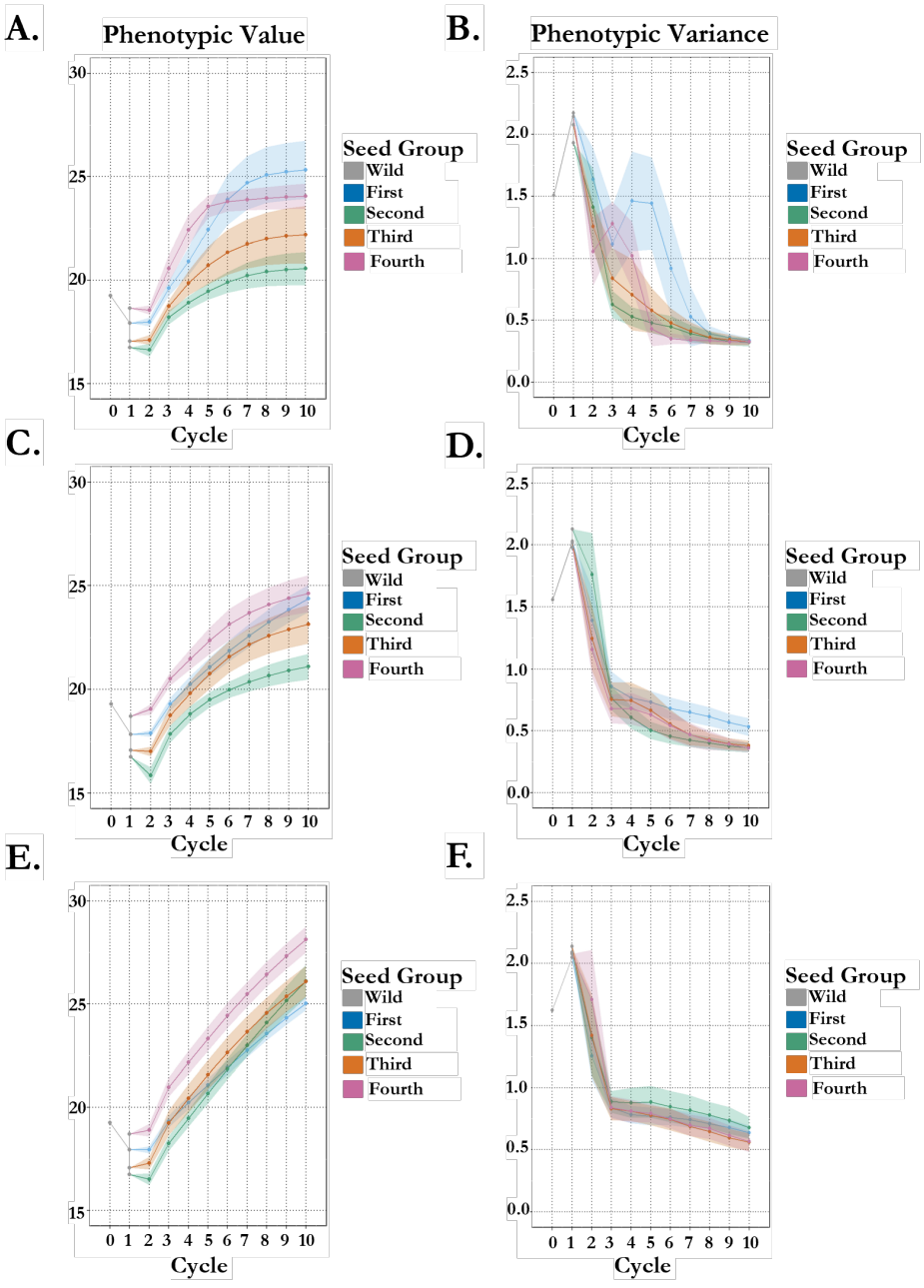
1.4.1 Trait associations and influences on gain of vigor

Seedling attributes such as height, root mass, abiotic and biotic resistances, and nutrient status are widely recognized to be critical components of plant success, especially in perennial tree species (Grossnickle and MacDonald, 2018). Improving genetic gain of vigor, measured as seedling height in our study, is especially important in forestry, where the cycle from planting to harvest is extended. Vigor can be selected at early stages and decrease the breeding cycle time as identified in *Acacia mearnsii* (Bisognin et al., 2023). Orchard stand and yield potential is directly impacted by early-stage vigor as well as stress resistance. This has been observed in the contemporary crops cotton and alfalfa. For example, more vigorous cotton seedlings have a shortened duration of sensitivity to pathogens by the reduction of fungal penetration into developed woody tissue (Bourland, 2019). There also is a slight negative correlation between fusarium wilt in alfalfa and vigor, indicating that selection for fusarium wilt resistance might increase vigor even in the absence of disease (Fonseca et al., 1999). Additionally, vigor is an important indicator of yield potential, as observed in wild soybean (Kofsky et al., 2020). These patterns are also observed in forestry, where in southern pine production, seedling quality and vigor play a critical role in survivability and growth potential (Johnson and Cline, 1991, South et al., 2001), also observed in resistant eco-types of koa (Dudley et al., 2015, 2020). The improvement of vigor, which is entangled with resistance, indicates the progress of population improvement during incipient domestication. The selection of FOXY resistant types and removal of low durability of resistant genotypes from the breeding population results in observed improvements of both traits in koa (Table 1 & 2). Our breeding population continues to generate more vigorous and resistant progeny through this selection regime. Moreover, simulation predicts the population to maintain this progress of seedling vigor improvement (Figure 5). However, simulation does identify optimal crossing parameters given our population and genetic architecture, where maintaining a larger breeding population and having fewer progeny per cross (Scheme 1; Figure 5A) will improve gain substantially over alternative schemes (Figure 5C/E). Each generation consists of more vigorous and resistant progeny, pushing the population towards a domesticated state which will successfully fill the cultural, economic, and ecological role of wild koa.

1.4.2 Bottleneck size, genetic load, and effects on LD

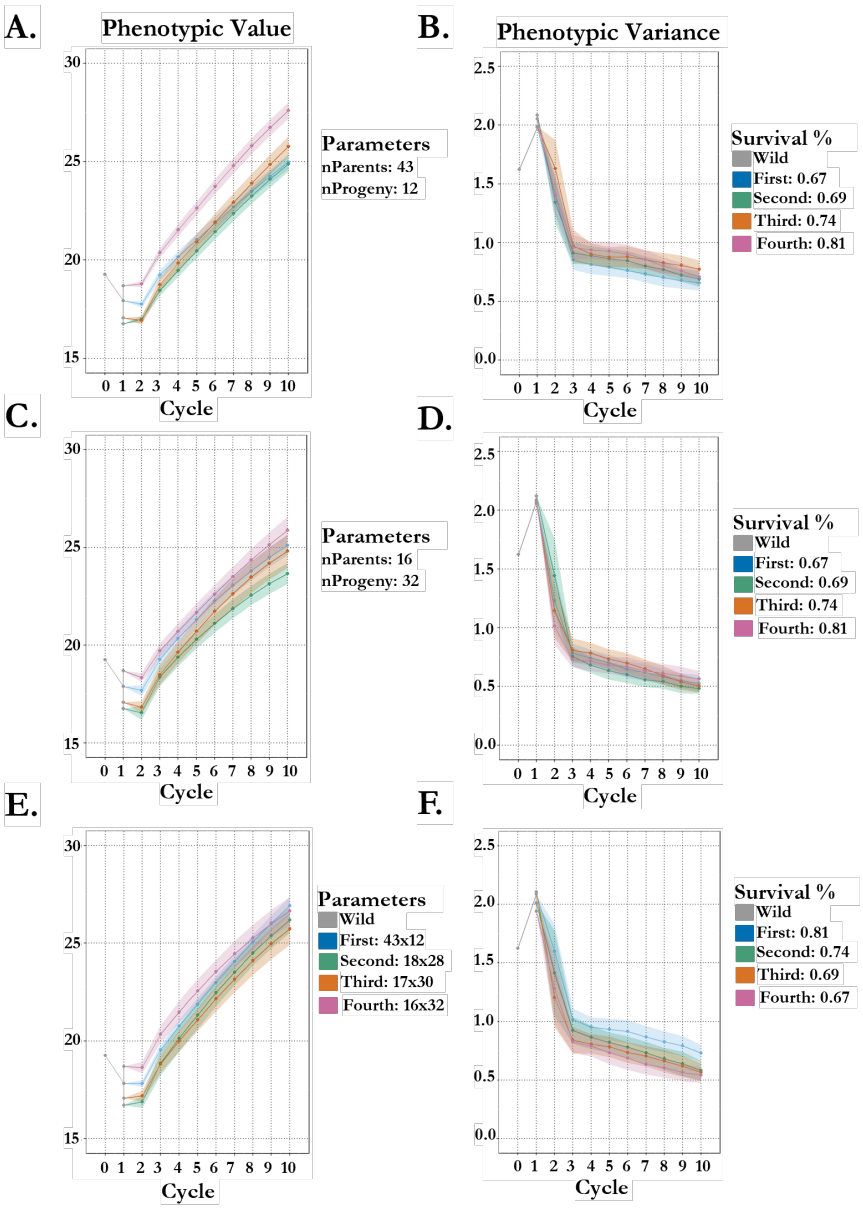
Genetic variation is an extremely important consideration during neo-domestication, where historical domestication incurred substantial bottlenecks (Eyre-Walker et al., 1998, Zhu et al., 2007, Allaby et al., 2008). Moreover, loss of variation during selection limits the potential for sustainable quantitative trait improvement (Lande, 1979). Our case study into

Figure 4: 100 stochastic simulations of seedling vigor (height in cm) in koa with different crossing parameters and survivability by seed group origin. Phenotypes are estimated in 1 environment. The plotted points and lines represent the mean of replications per cycle with the shaded regions representing the standard deviation of replications per cycle. (A/B) Genetic architecture of seedling vigor is simple oligogenic (8 QTL). (C/D) Genetic architecture of seedling vigor is complex oligogenic (20 QTL). (E/F) Genetic architecture of seedling vigor is polygenic (100 QTL). (A/C/E) represent the phenotypic gain, measured as phenotypic value and (B/D/F) represent the phenotypic variance.



koa domestication extends the important considerations to plasticity and adaptability. The

Figure 5: 100 stochastic simulations of seedling vigor (height in cm) in koa with altered schemes by different crossing parameters and survivability by seed group origin. Phenotypes are estimated in 1 environment and the genetic architecture of seedling vigor is simulated as polygenic (100 QTL). The plotted points and lines represent the mean of replications per cycle with the shaded regions representing the standard deviation of replications per cycle. (A/B) Scheme 1: seed group survival % is same as empirical estimates while number of parents and progeny are set constant, 43 and 12 respectively. (C/D) Scheme 2: seed group survival % is same as empirical estimates while number of parents and progeny are set constant, 16 and 32 respectively. (E/F) Scheme 3: seed group survival % is inverted from empirical while number of parents and progeny is same as empirical. (A/C/E) represent the phenotypic gain, measured as phenotypic value and (B/D/F) represent the phenotypic variance.



improved types are not only for the agroecosystem but also for reforestation efforts, where a narrow-genetic base could spell abiotic and biotic disaster considering the worsening climatic conditions and expanding global trade (Foley et al., 2011, Ramankutty et al., 2018, Chapman et al., 2017). As we begin to bottleneck our population during domestication, consideration of realized and masked genetic load is important. Bottlenecks purge some deleterious mutations (reducing load) but it also converts masked load into realized load with prolonged bottlenecks fixing deleterious mutations (Bertorelle et al., 2022). The ratchet effect is a common cost of domestication as regions of low effective recombination, often the selected haplotypes, disproportionately accumulate deleterious variants (Kono et al., 2016). Balance can be restored using migration (genetic rescue), a component worth consideration in breeding scheme development for neo-domestication. Our breeding schemes (Figure 4) highlight the influence of bottleneck size on genetic variation after 10 cycles of selection, where Scheme 1 (43 parents) possesses $\tilde{2}5\%$ more variation than Scheme 2 (16 parents). However, it should be expected that following incipient domestication effective population sizes become small, like crop species, creating a strong prevalence of genome-wide linkage disequilibrium that should validate the use of genomic selection (Voss-Fels et al., 2019). Careful consideration is necessary though because high LD decay is expected during the bottlenecking of an outbred species with high effective population size, reducing the predictive ability of genomic selection (Zhang et al., 2016).

1.5 Conclusion

In the koa example, large improvements are realized in a few cycles of selection even despite the long temporal time of this perennial hardwood tree species. Our initial selection on agronomic traits has empirically improved the adaptability of our populations to the agroecosystem and natural ecosystem. These improvements in adaptation to the agroecosystem highlight these agronomic traits as important domestication syndrome for koa. Furthermore, orchard stand is improved, meaning the breeding program can shift to quality traits such as rich and complex fiddleback grain. The stochastically simulated breeding schemes outline rapid domestication when traits of interest are known. Moreover, the speed of this change in phenotypic performance occurred regardless of the underlying genetic architecture, giving promise to the introduction of new crops with alternative domestication syndromes. Our empirical and simulated evidence of the trend towards domestication in *Acacia koa* outlines parametrization of crossing in the breeding cycle for rapid improvement of adaptation to the agroecosystem.

Chapter 2: *Theobroma cacao* variety differentiation with altered phenotyping precision - informing breeding cycle evaluation expectations

2.1 Introduction

Theobroma cacao L. (hereafter, cacao) is a domesticated tropical fruit crop essential to the chocolate industry. The center of diversity and origin of cacao is the border of Ecuador, Colombia, and Brazil with first reports of domesticated uses by the Mayan peoples of Mesoamerica (Clement et al., 2010, Thomas et al., 2012). Cultivated and wild cacao resemble one another, signifying there was, and continues to be, the identification of superb accessions from wild stands which were distributed for production (Laliberté et al., 2012). Hybrids generated from open-pollinated landraces are now major production varieties with disease resistance across the world. Therefore, cacao in production can range from wild to open pollinated landrace selections, to newly bred hybrids, where the cultivated types are no more adapted to the agroecosystem than their wild counterparts, despite reduction in allelic diversity in landraces (Motamayor et al., 2008, Laliberté et al., 2012). Despite this range of domestication states, most cacao production is based on landraces selected prior to 1950 with only one-third of cacao globally derived from hybridization and crop improvement (Laliberté et al., 2012). Wild genetic diversity is not represented in production varieties, but this diversity is available for introgression of important traits to improve productivity of the crop in the world's tropical agroecosystems (Bekele and Phillips-Mora, 2019).

Previous work has identified three groups of cacao based on phenotypic clustering (Cheesman, 1944, Lachenaud and Oliver, 2005). Amazon Forastero cacao is cultivated on 70% of cacao farms due to strong flavor, high butterfat content and disease resistance (Eskes and Lanaud, 2001, Iwaro et al., 2003, Khan et al., 2008). Criollo cacao has a limited production range to mostly Central America due to high disease susceptibility and displays low vigor, likely due to its low genetic diversity (Motamayor et al., 2002). Trinitario cacao is a highly heterogeneous group derived from the hybridization of Amazon Forastero and Criollo with high levels of phenotypic and genetic variability, a key in clonally (i.e., grafted) propagated crops (Lass and Wood, 1985, Motamayor et al., 2003). (Motamayor et al., 2008) found ten genetic clusters of cacao (study did not include Trinitario types) with a focus on wild accessions. This distinction is important for the maintenance of long-term adaptability and productivity in crops, however, selection of individual accessions from a large survey (∼1200 accessions) for use as parents is costly (i.e. land and labor).

Selection of a new cultivar in possession of a key trait is often done through accurate and precise phenotyping, even though marker-assisted selection is used in structured, more developed breeding programs. The phenotyping must be precise enough where estimated

differences between individuals are identified outside of confidence intervals while maintaining cost minimization. Moreover, the precision of phenotyping (sub-sampling) should be constrained to a set point where these differences are observable through data analysis and genotypic performance ranking. For example, in cotton heat tolerance research, the phenotypic correlation of quantity and quality traits by high-throughput phenotyping (HTP) and physiological hand collected phenotypes were mostly significantly correlated (Pauli et al., 2016). Furthermore, impactful gains in a developed program can be realized with more fine-tuned phenotyping (Cobb et al., 2013). Variation in the precision of phenotyping will often result in different ranks, therefore using the correct amount of phenotyping will assist in expediting selection of accessions and avoid situations such as the narrow genetic base of historic cacao varieties that resulted in disease susceptibility during the mid-1900's (e.g., witches' broom, frosty pod rot). The disease susceptibility necessitated the use of exotic germplasm usage in breeding programs to incorporate resistance, a process that has been slow to deliver new cultivars in the short-lived perennial. Rapid and appropriate phenotyping techniques permit efficient and accurate selection of exotic germplasm when it is of utmost importance.

Population improvement through artificial selection requires optimization of genetic gain for different traits (Gaynor et al., 2017). This is achieved through an iterative process of generational increase in favorable alleles in the population under selection, acting to increase the probability of extracting a superior cultivar from the population (Cobb et al., 2019, Van Tassel et al., 2020). The increase in favorable alleles drives genetic gain, which is a product of additive genetic variation within the population, selection intensity, and selection accuracy. Each gain component can be increased through different methods and technology applied across the breeding cycle (Hickey et al., 2017, Wallace et al., 2018, Wartha and Lorenz, 2021). Population improvement relies upon adequate and appropriate parametrization in the breeding cycle of crossing, evaluation, and selection (Covarrubias-Pazaran et al., 2022).

The focus of this study is the effect of changing precision of estimates during evaluation of diverse germplasm on the selection for and gain of important agronomic traits. Cacao is an excellent system to derive expectations because it provides the opportunity to understand how precision differentially affects selection across the domestication continuum of landrace and developed varieties. Therefore, the case study provides a clear set of experiments and evidence towards subsampling and replication procedures during evaluation of diverse germplasm. This is an opportunity for insight to develop expectations of gain during population improvement with variable evaluation precision to inform situational changes through breeding cycles in neo-domestication programs (Cobb et al., 2019, Covarrubias-Pazaran et al., 2022).

2.2 Materials and Methods

2.2.1 Germplasm and trials

The trial field (planted 2017) is a randomized complete block design with 15 treatments (genotypes-Table 3) ranging from relatively wild, open pollinated landrace variety selections (6 genotypes) to developed breeding lines (8 genotypes) and a control production variety (1 genotype) grafted to the same reliable rootstock variety to avoid rootstock-scion interaction

effects. There are 4 blocks and 15 treatments for a total of 60 plots (6 clones per plot, subsamples) and 360 total trees. The orchard site is in proximity to koa (*Acacia koa*) to the south, coffee (*Coffea arabica* and *C. canephora*) to the north, breadfruit (*Artocarpus altilis*) to the west, and mix forest to the east (Supplemental Figure 2).

Cacao has two distinct harvest seasons in Hawaii, the early winter and mid to late spring. The trees in this orchard site began pod production in winter 2020, the first harvest and data collection. There were 5 discrete harvests from the trial plot: winter 2020, spring 2021, winter 2021, spring 2022, winter 2022. The measurements are yield traits by tree (every sub-sample) including pod count, seed count/pod (3 pods), seed weight/pod (3 pods), total seed weight. Therefore, the average seed count and weight per pod, tree, or plot can be calculated and subsampling can be used to simulate differing phenotyping precision.

Table 3: List of cacao germplasm included in the trial planted in 2017 at Hawaii Agriculture Research Center – Maunawili Station.

Cultivar	Year released	Cultivar style	Breeding State	Origin
MTc10-02 (CCN-51 X 2057)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-03 (UNAP-2 X TIP-1)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-04 (SIL-1 X D-147)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-12 (AMAZ-14 X EBC -148)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-05 (CCN -51 X B-60)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-07 (CCN-51 X B-60)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-09 (EET -387 X A-645)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
MTc10-17 (TAP-3 X TIP -1)	2012-2013	Hybrid	Elite cultivar	Mars, Inc.
ICS 95	1979	Unknown	Production	Trinidad
K25 OP	N/A	Open pollinated	Landrace	HARC, Hawaii
W1010709	N/A	Open pollinated	Landrace	Waialua, Hawaii
W1010202	N/A	Open pollinated	Landrace	Waialua, Hawaii
W1060205	N/A	Open pollinated	Landrace	Waialua, Hawaii
TARS 2	N/A	Open pollinated	Landrace	Puerto Rico
TARS 9	N/A	Open pollinated	Landrace	Puerto Rico

2.2.2 Analysis

To quantify the amount of phenotyping precision needed to observe significant differences between genotypes within groups, we fit three mixed-models. Data used in models is augmented differently to simulate alternative phenotyping schemes: max sampling (six trees/plot), half subsampling (random three trees/plot), and no subsampling (random one tree/plot). The mixed-models follow general form of total seed weight as response with random predictor block and fixed predictors of genotype and harvest date. This process is repeated with response mean seed size (weight as grams). Comparison of estimated marginal means of genotypes within each augmented precision model is used to identify rank change within and between variety groups (i.e. domestication continuum landrace and hybrid) and identify means outside of baseline model confidence intervals (0.95). Moreover, we fit another model form using mean seed size as the response with fixed effect continuum group and random effects genotype and harvest date. This is a specific change aimed to generate estimates for the domestication continuum groups and residual error, while removing variance attributable to genotype and harvest date.

2.2.3 Simulation

Once mixed-model analysis estimates group effect and predicts residual errors across variable sampling precision, we move to integrate to stochastic simulation using the R Statistical Software package 'AlphaSimR' (Gaynor et al., 2021) for future projection under variable trait genetic complexity. Model estimates and rank changes are then used to integrate stochastic simulation. Founding population parameters match estimates of our different continuum groups for mean seed size to parse gain in these groups given differential estimates and variability. For future projection, we simulate alternative genetic complexities for mean seed size as simple oligogenic (8 QTL) and complex oligogenic (20 QTL) over 10 cycles of selection. Founders are formed from species specific information such as 10 chromosomes in cacao and specifying mean and variance for the trait within each group (Seed Size; Table 4). Once the population is specified under a given genetic architecture, phenotypes are estimated by taking into account the additive genetic architecture with the residual error, estimated through our augmented precision mixed-models (Seed Size; Table 4). We then specify factors relevant to the breeding cycle, such as number of parents, number of crosses per parent, and the number of progeny per cross. The founding population therefore matches each seed groups' linear mixed-model estimates, forming the substrate of the simulation, based on truncation selection. However, truncation selection is augmented through the variable precision, interpreted, and employed by rank change. Max sampling serves as baseline and truncation selection of the top 33% of individuals is performed to serve as the parents for the following generation. Precision begins to deteriorate when 50% subsampling is performed, where landrace remains as the baseline (no rank change among genotypes) but hybrid derived varieties have 50% of selections from top 33% performers and 50% of selections from 34-66% performers. Precision does not deteriorate further in hybrids under no subsampling, but in landrace precision follows hybrid under 50% subsampling. Concisely, full subsampling is baseline, 50% subsampling affects the selection unit (individuals selected) in hybrids but not landrace, and no subsampling effects the selection unit (individuals selected) in hybrids and landrace. Reductions in precision result in imprecise selection of individuals which will have

direct effects on the rate of gain.

Next, we investigate the effects of increasing environmental evaluation under these reductions to precision of estimates. To do so, we select complex oligogenic (20 QTL) to serve as the genetic architecture of mean seed size, with the founding population of each continuum group being formed as previously mentioned. Therefore, we set 3 comparative schemes each with a gradient of environmental evaluation: (1) continuum groups under full subsampling precision each with 1, 3, and 5 environmental replications of evaluation; (2) continuum groups under 50% subsampling precision each with 1, 3, and 5 environmental replications of evaluation; and (3) continuum groups under no subsampling precision each with 1, 3, and 5 environmental replications of evaluation. These schematic alterations will inform our understanding of the effects of environmental replication at alleviating the effects of reduced subsampling. We define our discrete environments as location-by-year, which provides the environmental variance-covariance metric integrated for simulation.

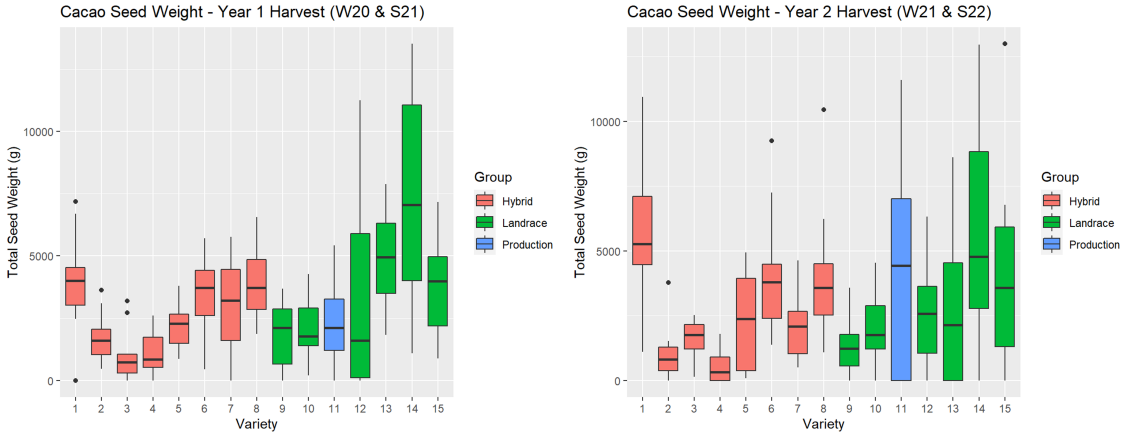
2.3 Results

2.3.1 Cultivar estimates and their rank change along a precision gradient

Yield of cacao varieties varied within and between groups, with landrace types exhibiting larger variation (Figure 6). Certain varieties within each of these groups (Hybrid – 1/MTc10-12 and Landrace - 14/W1060205) maintain the mean of total seed weight greater than the production control mean through the two harvest seasons (Figure 6). Mixed-model analysis of variety total seed weight with varied precision (altered sampling) produces varied levels of clarity. Full sub-sampling (Figure 7A) exhibits distinction within landrace variety type groups by estimated marginal means with 0.95 confidence intervals, such as W1060205 being better than K25 and W1010202, which in turn are better than TARS 2. However, differences between hybrid varieties requires expansion past the next LSD group (Supplemental Figure 3). Decreasing precision of estimates through reduced subsampling reduced the clarity of differences to the point of being unable to observe statistically significant differences between varieties from the landrace group as well as between varieties from the hybrid group (Figure 7C). The highest yielding variety through all precision levels is W1060205 (landrace), highlighting the variety's performance through all subsamples (trees). The next best variety is MTc10-12 (hybrid) when subsampling is used, but the loss of subsampling decreases this variety's rank from 2 to 4, highlighting this variety's lack of consistent tree performance and the importance of subsampling for precise selection. Rank change (5 of 15) is restricted to only 2 levels (increasing or decreasing) when reducing precision of estimates from full to 50% subsamples (Figure 7A/B). However, ranks change at higher frequency (12 of 15) and levels (6) when precision of estimates is reduced from 50% subsamples to a single sample (Figure 7B/C).

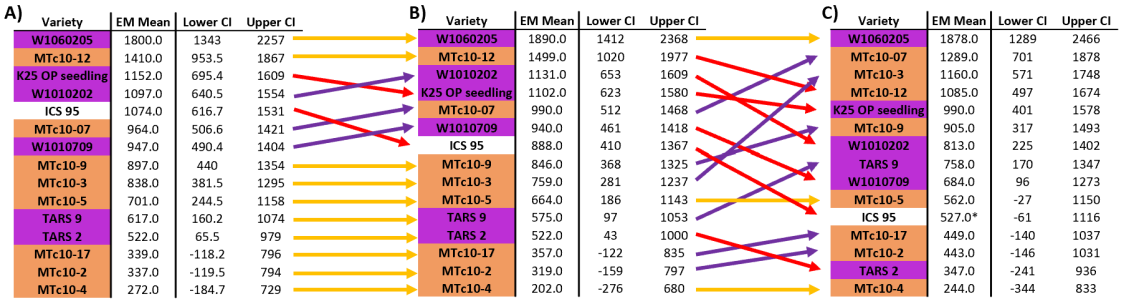
When the trait of interest is changed to mean seed size, an important attribute for roasting during the chocolate making process, the best varieties change (Figure 8). MTc10-12 is the best variety for mean seed size, only decreasing by one level in rank when 50% of samples are used (Figure 8). W1060205 was identified as the highest yielding variety by total seed weight, achieved through small seed size with rank of <11 in all sampling schemes (Figure 8).

Figure 6: Total seed weight variability through sub-harvests in A) year 1 (winter 2020 and spring 2021) and B) year 2 (winter 2021 and spring 2022).



Seed size is a trait with little clarity of differentiation (three LSD groups in full sub-sampling and 1 in single sample; Supplemental Figure 4) and higher frequency of rank change during altered phenotyping precision (11 of 15 Figure 8A/B and 13 of 15 Figure 8B/C).

Figure 7: Estimated marginal mean rank change across varied sub-sampling for total seed weight (grams) with A) full sub-sampling, B) half sub-sampling, and C) single sample. Arrows represent no rank change (yellow), negative rank change (red), and positive rank change (purple). Varieties a marked as open-pollinated landrace (magenta), hybrid variety (orange), and production control (none). Asterisks next to EM mean represent those EM means of reduced subsampling outside of the confidence intervals of baseline.



A)

Variety	EM Mean	Lower CI	Upper CI
W1060205	1800.0	1343	2257
MTc10-12	1410.0	953.5	1867
K25 OP seedling	1152.0	695.4	1609
W1010202	1097.0	640.5	1554
ICS 95	1074.0	616.7	1531
MTc10-07	964.0	506.6	1421
W1010709	947.0	490.4	1404
MTc10-9	897.0	440	1354
MTc10-3	838.0	381.5	1295
MTc10-5	701.0	244.5	1158
TARS 9	617.0	160.2	1074
TARS 2	522.0	65.5	979
MTc10-17	339.0	-118.2	796
MTc10-2	337.0	-119.5	794
MTc10-4	272.0	-184.7	729

B)

Variety	EM Mean	Lower CI	Upper CI
W1060205	1890.0	1412	2368
MTc10-12	1499.0	1020	1977
W1010202	1131.0	653	1609
K25 OP seedling	1102.0	623	1580
MTc10-07	990.0	512	1468
W1010709	940.0	461	1418
ICS 95	888.0	410	1367
MTc10-9	846.0	368	1325
MTc10-3	759.0	281	1237
MTc10-5	664.0	186	1143
TARS 9	575.0	97	1053
TARS 2	522.0	43	1000
MTc10-17	357.0	-122	835
MTc10-2	319.0	-159	797
MTc10-4	202.0	-276	680

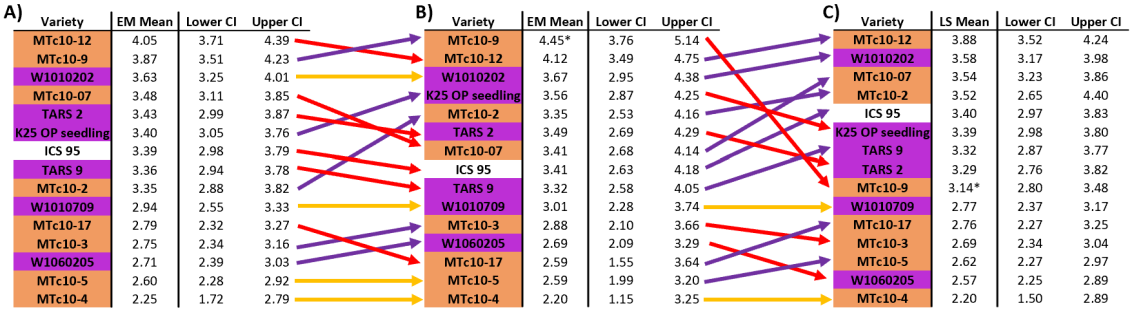
C)

Variety	EM Mean	Lower CI	Upper CI
W1060205	1878.0	1289	2466
MTc10-07	1289.0	701	1878
MTc10-3	1160.0	571	1748
MTc10-12	1085.0	497	1674
K25 OP seedling	990.0	401	1578
MTc10-9	905.0	317	1493
W1010202	813.0	225	1402
TARS 9	758.0	170	1347
W1010709	684.0	96	1273
MTc10-5	562.0	-27	1150
ICS 95	527.0*	-61	1116
MTc10-17	449.0	-140	1037
MTc10-2	443.0	-146	1031
TARS 2	347.0	-241	936
MTc10-4	244.0	-344	833

2.3.2 Yield component performance of domestication continuum groups

Following the identification of cultivars within landrace and hybrid groups with high yield and mean seed size along a precision gradient, we assign another linear mixed-model for estimation of group parameters. The four yield component traits of interest are mean seed size (grams), total seed weight (kilograms), number of pods per tree (count), and number of seeds per pod (count). Mixed-model analysis finds a common gradient where production estimates are greater than landrace estimates, which in turn are greater than hybrid estimates

Figure 8: Estimated marginal mean rank change across varied sub-sampling for mean seed size (grams) with A) full sub-sampling, B) half sub-sampling, and C) single sample. Arrows represent no rank change (yellow), negative rank change (red), and positive rank change (purple). Varieties a marked as open-pollinated landrace (magenta), hybrid variety (orange), and production control (none). Asterisks next to EM mean represent those EM means of reduced subsampling outside of the confidence intervals of baseline.



(Table 4). The standard error of the estimates follows this same gradient. However, the number of seeds per pod is the only trait where the estimate gradient places production below hybrid with landrace having the largest (Table 4). Although the production group estimates are based on a single production variety (ICN 95) via replication, total seed weight is not degraded due to a large mean seed size.

Table 4: Linear mixed-model estimates of yield component traits across different domestication continuum groups using max sampling.

		Mean Seed Size (g)	Total Seed Weight (kg)	Pod/Tree (#)	Seeds/Pod (#)
Group Estimate	Hybrid	2.99 (0.19)	1.25 (0.018)	9.96 (1.32)	73.07 (9.19)
	Landrace	3.17 (0.25)	1.67 (0.025)	12.86 (2.02)	82.26 (9.16)
	Production	3.37 (0.49)	2.03 (0.050)	15.61 (3.97)	64.49 (18.04)
Var-Comps	Genotype	0.21 (0.46)	196.55 (0.044)	12.34 (3.51)	265.6 (16.30)
	Harvest	0.04 (0.19)	24.15 (0.016)	0.00 (0.00)	241.9 (15.55)
	Residual	0.44 (0.66)	1470.82 (1.213)	103.32 (10.17)	1,432.1 (37.84)

2.3.3 Stochastic simulation for projected seed size gain along varied evaluation precision

Stochastic simulation was used to project gain in mean seed size across a precision gradient. We begin with parameters matching empirical estimates for both of our cultivar groups (landrace and hybrid) to project gain in mean seed size through 10 cycles of selection with varying genetic architecture complexity, taking into account the effect of reduced precision of estimates. Simple oligogenic control (8 QTL) of mean seed size finds a 42% gain over 10 cycles of selection with baseline precision and hybrid cultivar founders (Figure 9A). The reduction to 50% subsampling of hybrid cultivars decreases gain over 10 cycles of selection to 38% (Figure

9C). Complex oligogenic control (20 QTL) of mean seed size finds a 73% gain over 10 cycles of selection with baseline precision with hybrid cultivars (Figure 9A). The reduction in gain increases with genetic complexity, where 50% subsampling reduces projected mean seed size gain to 59% over 10 cycles. Interestingly, the effect of reducing subsampling (precision) is smaller on the landrace group, even after considering that the effect is not noticed until no subsampling is conducted. Under simple oligogenic control mean seed size has projected gain of 46%, only decreasing to 45.5% gain (Figure 9A/E). Increasing genetic complexity for the trait does result in a larger reduction to gain (67% to 57%), but still less than half the loss realized by the hybrid group (Figure 9). Despite this reduction of gain in both continuum groups when precision decreases, trait variation decreases at the same rate (Figure 9B/D/F). The different gradient of reduced precision, 50% subsampling in hybrid and no subsampling in landrace, could be alleviated through increased environmental replication.

We therefore implement three different levels of environmental replication within each precision level, maintaining 1 evaluation environment as the baseline, increasing to 3 and 5 evaluation environments on a complex oligogenic trait (20 QTL) from landrace and hybrid cultivar estimates. Under full subsampling, the rate of gain from hybrids in the complex oligogenic trait (mean seed size) over 10 cycles of selection is 74%, 41%, and 54% using 1, 3, and 5 evaluation environments, respectively (Figure 10A). However, when precision of estimates is decreased through reduction of subsampling to 50% and none, rate of gain over 10 cycles of selection is simulated at 53%, 33%, and 47% (Figure 10C/E). A shift of cultivar group to landrace under full subsampling finds simulated gain over 10 cycles of selection of 81%, 93%, and 98% under 1, 3, and 5 environments, respectively (Figure 10A/C). These gains are expected under 50% subsampling as well because rank change of selection does not occur until no subsampling. Under no subsampling, the simulated gain over 10 cycles of selection is 72%, 65%, and 57% (Figure 10E). Population phenotypic gain for a complex oligogenic trait, representing mean seed size, is larger using landrace cultivars as the founders by about 10% greater than hybrid cultivar founders using full subsampling (Figure 10). Moreover, when reductions to subsampling is considered, landrace cultivars are more tolerant to the loss of precision, not losing gain until no subsampling occurs. The proportion of loss in genetic gain is also different between populations formed from hybrid and landrace founders, where only 12%, 5%, and 8% differences in genetic gain are found in landrace while 22%, 8%, and 7% differences in hybrid under 1, 3, and 5 environments, respectively (Figure 10A/C/E).

Stochastic simulation of increased evaluation environments does find heightened variation via reduced measurement error. For example, using the founding population through landraces predicts loss of variation (Figure 10B/D/F). However, results are also predicted over 10 cycles of selection using hybrid founders where variation increases (Figure 10B/D/F). The loss of variation in the landrace founding population across schemes is occurring due to a greater available genetic variance present over the predicted residual error.

2.4 Discussion

2.4.1 Interspecies comparisons of seed size change

Figure 9: 100 stochastic simulations of mean seed size (g) in cacao using founding populations from different continuum groups and alternative genetic complexity. Phenotypes are estimated in 1 environment. Genetic architecture of mean seed size is simple (8 QTL) or complex oligogenic (20 QTL) with estimates for founding population derived from hybrid (H) or landrace (LR) populations. The plotted points and lines represent the mean of replications per cycle with the shaded regions representing the standard deviation of replications per cycle. (A/B) Baseline model using full subsampling and 30% truncation selection of highest performing individuals. (C/D) Reduction to precision model using 50% subsampling and 30% truncation selection of highest performing individuals. (E/F) Reduction to precision model using no subsampling and 30% truncation selection of highest performing individuals. (A/C/E) represent the phenotypic gain, measured as phenotypic value and (B/D/F) represent the phenotypic variance.

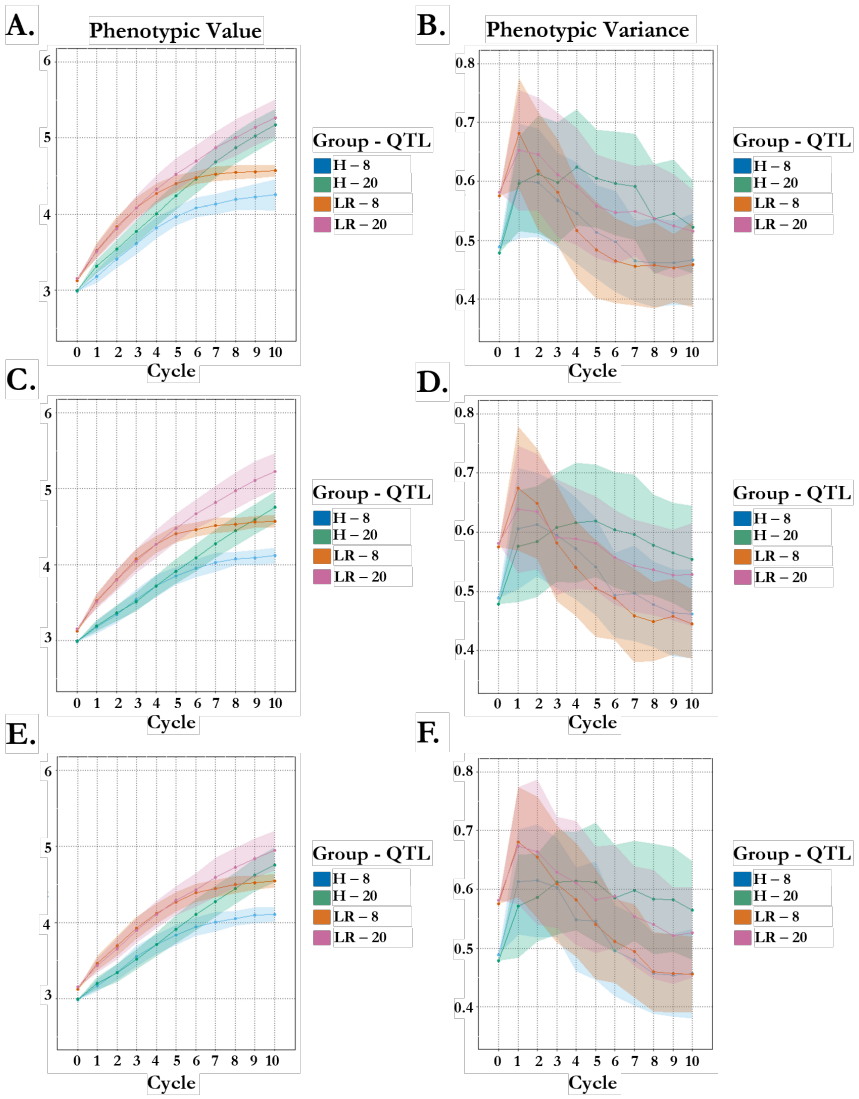
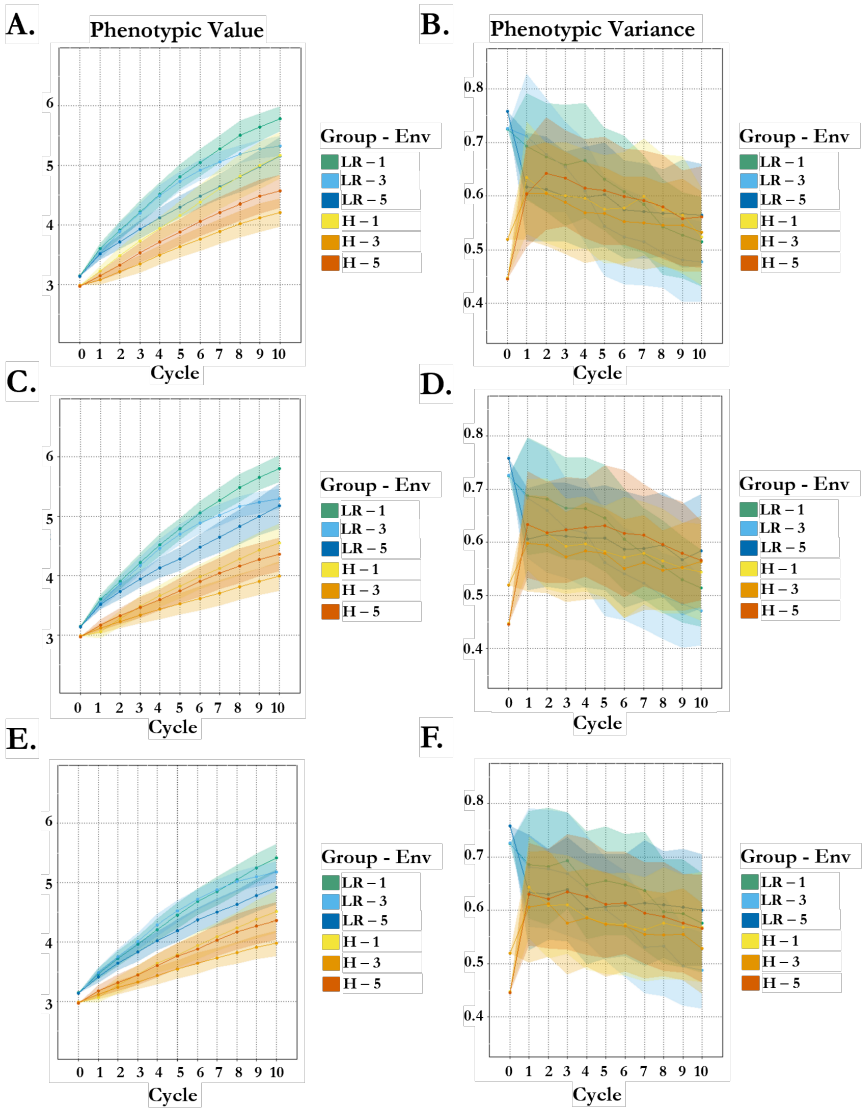


Figure 10: 100 stochastic simulations of mean seed size (g) in cacao using founding populations from different continuum groups and alternative genetic complexity. Phenotypes are estimated in 1, 3, and 5 environments. Genetic architecture of mean seed size is simple (8 QTL) or complex oligogenic (20 QTL) with estimates for founding population derived from hybrid (H) or landrace (LR) populations. The plotted points and lines represent the mean of replications per cycle with the shaded regions representing the standard deviation of replications per cycle. (A/B) Baseline model using full subsampling and 30% truncation selection of highest performing individuals. (C/D) Reduction to precision model using 50% subsampling and 30% truncation selection of highest performing individuals. (E/F) Reduction to precision model using no subsampling and 30% truncation selection of highest performing individuals. (A/C/E) represent the phenotypic gain, measured as phenotypic value and (B/D/F) represent the phenotypic variance.



The specific trait of interest in our study is seed size, a well-studied domestication syndrome component to yield and agronomic performance, such as early vigor and planting depths (Purugganan and Fuller, 2009, Sedbrook et al., 2014). Variation is found within and between our continuum groups, with lower phenotypic variation found in the narrow genetic base of hybrid types (Table 4), also found in new world bean cultivars following historical domestication (Kaplan, 1965). Our use of stochastic simulation predicts phenotypic gains in our *in-silico* trait representing mean seed size following directional selection for the trait (Figure 9). Findings in other neo-domestication projects empirically observe similar gains over few cycles of selection in intermediate wheatgrass and silphium (DeHaan et al., 2018, Vilela et al., 2018). This suggests that the response to selection for this trait is in fact as great as simulated, despite the route of improvement in historical domestication argued to be at least a mixture of unconscious and conscious selection (Fuller, 2007). It is important to note that gain in this trait, like any other, is restricted by available variation in which selection can be performed (Cobb et al., 2019). Adequate collections of wild germplasm, avoidance mating, and/or genetic rescue/migration during population improvement in neo-domestication will likely play major roles towards later stages. These strategies will require conscious implementation, whereas under historical domestication they have been postulated to be unconsciously performed (Smỳkal et al., 2018). Lastly, seed is a reproductive organ used for consumption and increases to alternative reproductive organs such as cultivated tubers of sweet potato and cassava as well as phytoliths in bananas/plantains have been found to be markedly larger than wild counterparts (Fuller et al., 2014). Accordingly, selection for increases in size to sexual/asexual and aboveground/belowground reproductive organs under historical and incipient domestication is possible, yet its sustainability will rely on variation and germplasm resources (von Wettberg et al., 2020).

2.4.2 Precision and/or accuracy during neodomestication

Proper characterization of phenotypic variation is heavily reliant upon experimental design and statistical analysis. Our study helps to elucidate some of the effects to selection accuracy under alternative estimation precision across different continuum groups (Figure 9) as well as the lessening of the environmental effect through increased environments of evaluation (Figure 10). Accuracy and precision are related during the plant breeding process where selection accuracy relies on the adequate precision of estimates (Resende and Alves, 2022). Plant breeding relies on the characterization of genotypic performance to identify and select the best, thereby maximizing genetic gain. Therefore, the correct ranking of genotypes and the precision of estimating their differences is highly relevant to cultivar improvement (Schmidt et al., 2019). Precision mainly depends on the number of observations related to a genotype and the structure of the design, where more information in balanced design improves precision (Laloë, 1993). But precision of estimates has many potential routes for improvement, including the improvement to trial design, the incorporation of genetic relationships among varieties, and the leveraging of analysis of covariance through correlated component traits (Mackay et al., 2015). These estimations are important towards accurately defining heritability for the given trait of interest, a component directly affecting response to selection or genetic gain. The selection of a trait with a high narrow sense heritability has the potential for more rapid improvement than those with low heritability, even under

less intensive evaluation schemes. Moreover, increases in replication and environment during evaluation will increase narrow sense heritability (Dudley and Moll, 1969). This is the case, in part because the importance and impact of subsampling on the precision of estimates is usually proportional to the level of stability in the genotype (Becker and Leon, 1988). The stability in a genotype is an essential concern during the phenotypic evaluation of different genotypes and expanding the environments of evaluation to cover the target population of environments (including those inhibiting crop success) is key to describing variation for the trait of interest and to accurately define adaptable and stable performance (Finlay and Wilkinson, 1963). However, this expansion of evaluation environments is logistically and financially constrained, especially in neo-domestication projects with limited market share and funding resources. Appropriate strategies towards resource allocation are therefore important and should be further investigated and developed (Lorenz, 2013). Nothing comes without a cost, hedging reductions of gain due to reduced precision of estimates (less subsampling) by increasing measurement accuracy (more evaluation environments) will more rapidly bottleneck the population. Therefore, careful consideration during strategic planning for evaluation is important and should be aimed towards specific goals for specific traits of interest with particular focus on genetic complexity and the status along the domestication continuum.

Chapter 3: *Stevia rebaudiana* phenotypic recurrent selection improves population flowering response - informing breeding cycle selection expectations

3.1 Introduction

Stevia rebaudiana Bertoni (hereafter, stevia) is a perennial, herbaceous shrub native to Paraguay which accumulates steviol glycosides in its leaves (Brandle et al., 1998). Historically the plant has been used as general sweetening agent by the indigenous Guarani peoples of Paraguay and Brazil (Soejarto, 2001), but it has only been cultivated as a global crop since the latter half of the 20th century (Lewis, 1992). Today, the crop is grown on every continent (Midmore and Rank, 2002). Global expansion of the crop highlights its strength as a non-caloric substitute in the natural sweeteners market (Kinghorn et al., 1984, Matsui et al., 1996), as well as medicinal applications due to its high level of bioactive compounds (Clemente et al., 2021).

Stevia is clonally propagated crop, due to the ease of cloning and the variable and unpredictable trait values resulting from the near obligate outcrossing mating system. Paralleling the expansion of production is the increase in the number of cultivars released, at the last report there were 90 commercially available (Angelini et al., 2018). These cultivars are typically referred to as non-adaptive and low-value, as they were derived through open-pollinated mass selection from few seminal parents (e.g., "Eirete", "Criolla", and "Morita") with selection focused primarily on steviol glycoside content (Brandle et al., 1998, Yadav et al., 2011, Cosson et al., 2019). This has led to most other traits of interest (TOIs) being ignored even though they represent many classic domestication traits including reduced seed dormancy, photoperiod sensitivity, branching, along with the important traits for all commercial species such as yield, biotic stress tolerance and abiotic stress resistances. These traits have different types of genetic control ranging from single gene to highly quantitative (Meyer and Purugganan, 2013). Improvement of these traits in turn improves agroecosystem adaptation and agronomic performance (Jungers et al., 2023).

The short history of production, even shorter history of crop improvement and lack of genetic/genomic resources has produced a situation where this crop remains semi-adapted to common agroecosystems. When a cultivated plant is less adapted to the agroecosystem compared to wild collected plants is a key determination of whether that crop is fully domestic (Harlan et al., 1975). Selection drives adaptation to the agroecosystem, the first step is usually through the selection of a suite of traits classically termed 'the domestication syndrome' (Harlan et al., 1973), which facilitate agronomic tasks. Stevia lacks the improvement of many of classic domestication syndrome traits including reduced dormancy, branching, and photoperiod sensitivity (Brandle et al., 1998, Yadav et al., 2011, Cosson et al., 2019). The improvement of traits requires variability within breeding populations (Lynch et al., 1998,

Bernardo, 2002). Despite the lack of improvement for many of these domestication traits in stevia, there has been variability documented in wild populations (Valio and Rocha, 1977, Zaidan et al., 1980) indicating the capacity for improvement of syndrome traits (Lee et al., 1982, Brandle and Rosa, 1992, Shyu et al., 1994, Shizhen, 1995). Therefore, appropriate selection and breeding techniques should be able to shift domestication syndrome trait values within breeding populations.

Population improvement through artificial selection requires optimization of genetic gain for different traits and tailored towards the end goal (Gaynor et al., 2017). This is achieved through an iterative process of generational increase in favorable alleles in the population under selection, acting to increase the probability of extracting a superior cultivar from the population (Cobb et al., 2019, Van Tassel et al., 2020). The increase in favorable alleles drives genetic gain, which is a product of additive genetic variation within the population, selection intensity, and selection accuracy (Figure 1). Each gain component can be increased through different methods and technology applied across the breeding cycle (Hickey et al., 2017, Wallace et al., 2018, Wartha and Lorenz, 2021). Population improvement relies upon adequate parametrization in the breeding cycle of crossing, evaluation, and selection (Covarrubias-Pazaran et al., 2022). Each step in a breeding cycle has multiple parameters and creating a global optimum may result in less-than-optimal improvement for any given trait.

The focus of this chapter is the effect of selection on an early generation semi-domesticate, where potential decisions include the percentage of individuals selected (selection intensity), selection method (culling, index, tandem), the selection unit (family, line), and the selection criteria (phenotype, genotype, breeding value). This chapter provides general knowledge on how to begin breeding scheme development for wild and semi-wild breeding (neo-domestication), as stevia is an excellent model system to understand selection during incipient domestication, providing an opportunity to develop expectations of population improvement through phenotypic recurrent selection to inform selection in the breeding cycle in neo-domestication programs (Cobb et al., 2019, Covarrubias-Pazaran et al., 2022).

3.2 Materials and Methods

3.2.1 Germplasm and Breeding

3.2.1.1 Original Breeding Material

Germplasm was originally obtained by Jim Brandle and subsequently purchased by Sweet Green Fields, Inc. (SGF). Germplasm was planted in field trials in California during 2015 and in 2016, true seed was sent from California by SGF and planted at three field sites in Zhejiang, China. Approximately 9,000 plants were dug up from one site in China and moved to a field at the SGF Ningbo location. These 9,000 seedlings were not selected for any traits, so should be relatively representative of the population genetics from the original California field, given no selection of maternal parent nor pollen control. These 9,000 individuals were allowed to open pollinate and seed was collected at random and bulked together. This seed was then germinated, and individual plants were dug up and transplanted into greenhouses. These plants were allowed to grow for 1-2 years and selections were made from individual seedlings based on steviol glycoside content. This is the origin of the eight progenitor lines for

the Hawaii Agriculture Research Center (HARC) stevia breeding and improvement program. Given intellectual property constraints, these 8 lines will be referred to as A, B, C, D, E, F, G and H. Following the arrival in Hawaii, progenitor lines were allowed to openly cross in a randomized complete block to begin a pedigree mass selection program. Therefore, maternal lineage is documented, and seedlings are grouped into half-sib families. Equal amounts of seed were collected from each progenitor (1,500 each or 12,000 total) forming the inter-mated 1 (IM_1) generation.

3.2.1.2 First Breeding Cycle Selection

IM_1 seedlings were germinated in a greenhouse facility located in Kunia, Hawaii, 500 seeds from each half-sib family were selected for rapid germination and early vigor and transplanted to $125cm^3$ pots. The resulting 4,000 seedlings were placed under artificial light to extend the day-length photoperiod to 16 hours with incremental reduction by 0.5 hour every two weeks. Seedling culling was conducted when flowering was observed, resulting in an IM_1' population of 75 individuals per half-sib family (600 total) selected for lower photoperiod sensitivity. Therefore, IM_1' is derived through observational selection for germination, early vigor, and photoperiod sensitivity. These individuals were allowed to randomly mate through polycross and no pollen control, meaning each IM_1' has equal probability of serving as the paternal parent, to form the IM_2 generation. Equal amounts of seed (20-30) were collected from each IM_1' plant (serving as the maternal source of genetic variation) to form IM_2 generation of 1,500 per progenitor or 12,000 total.

3.2.1.3 Second Breeding Cycle Selection

The process of early selection in IM_2 generation followed the same protocol as the previous generation for germination and early vigor. However, an important caveat is that because only 20-30 individual seeds share the same maternal parent, and each family had uneven germination similar to previous reports in stevia (Brandle et al., 1998), some IM_1' maternal plants possess no representation in IM_2, even though original parental progenitors (A, B, C, D, E, F, G, H) maintain equal representation in the population (IM_2: 12.5% A-H). Furthermore, two augmented randomized complete block trials in Maunawili (MW) and Kunia, Hawaii with 700 unique varieties between them were conducted during the summer and fall 2020. The field trials possess 10 blocks with 35 plots per block for a total of 350 plots per field. Checks to account for within and between environment variation were 4 (MW: 19-0175 , 19-0241, 19-0091, 19-0407 & Kunia: 19-0107, 19-0121, 19-0172, 19-0085) and 3 (B, D, 19-0189) genotypes, respectively. The genotypes used in the field trials include progenitor lines, IM_1', and IM_2'. Parental selection of elite genotypes from IM_1' and IM_2' were identified via least-significant difference for photoperiod sensitivity, selecting those individuals with statistically different (lowered photoperiod) groupings from commercial checks. Elite genotypes were randomized into complete block design, consisting of 37 individuals (IM_{E1} =8 and IM_{E2} = 29) to form IM3. This crossing is the first in the program that does not equally represent each original progenitor; where A=10, B=1, C=0, D=10, E=6, F=6, G=1, and H=3. Approximately 300 seeds were collected from each individual for a total of 11,100 seeds in IM_3.

3.2.1.4 Third Breeding Cycle Selection

The same early selection protocol was followed in the IM_3 as in previous generations resulting in 2,000 plants that were subsequently tested in greenhouse conditions with day length extension. This second filter of IM_3 provided the material needed to estimate the breeding value of 20 individual inter-mated elite 1 (IM_{E1}) and 2 (IM_{E2}) varieties, where 12 randomly selected IM_3 seedlings were chosen for the first controlled environment trial in a shipping container with GE Arize Element L1000 PKB fixtures, air conditioning, and automatic irrigation. The elite varieties are from progenitors A (3), D (7), E (3), F (6), and H (1). Elite selections of IM_3 plants (IM_{E3}) from this trial were sourced from elite maternal lineage with the highest best linear unbiased prediction (BLUP) value for biomass, photoperiod, and TSG content.

3.2.2 Trials to understand genetic gain between cycles of selection

To understand the phenotypic gain and assess the breeding program improvements of traits, a second controlled environment trial was initiated as a randomized complete block with 4 blocks and 104 plots including all original progenitors (A-H) and varieties from every generation and lineage combination (prioritizing elite selections). The trait data collected includes plant height, branching, leaf width and length, internode spacing, TSG content, and photoperiod sensitivity. It is important to note that this is the first-time selection was rigorously made including steviol glycosides content to confirm quality maintenance in the program.

3.2.3 Analysis of gain per breeding cycle

To quantify differences in trait value and variation between stevia cycles of selection for different TOIs and parse these differences between generations, we fit two mixed-models: (1) with the form of TOI as the response with random predictors genotype (with additive kinship matrix), block, and cycle of selection with heterogeneous variance to account for varying genotypic variation in each cycle due to uneven sample sizes and quantify the shift in variation through generations for best linear unbiased prediction (BLUP) of genotype performance; and (2) with the form of TOI as the response with random predictors genotype (with additive kinship matrix) and block with homogeneous variance with cycle of selection as fixed to quantify genetic gain and generate best linear unbiased estimates (BLUE). Code is available on GitHub. Mixed-models performed with sommer (Covarrubias-Pazaran, 2016). Estimates of narrow-sense heritability are derived from the mixed-model with homogeneous variance component outputs. This follows form of additive genetic variance (generation plus variety effect) over the total genetic variance (additive genetic variance plus residual error).

3.2.4 Simulation integration and comparison

Once mixed-model estimates of generational effect and residual errors are generated, they can be integrated into a stochastic simulation (Gaynor et al., 2021) to generate expectations for future generations, estimating trait genetic architecture, as well as test unused, potential methodologies to improve gain and/or variance. For projections, simulation for alternative

genetic complexities for photoperiod causing flowering as monogenic (1 QTL), oligogenic (20 QTL), and polygenic (100 QTL) over 10 cycles of selection. Designing the founders takes species specific information such as 11 chromosomes in stevia as well as specifying mean and variance for the trait (progenitor generation, Table 5). Once a trait is specified under a given genetic architecture, phenotypes are estimated by taking into account the additive genetic architecture with the residual error, estimated through mixed-model (Table 5). Relevant factors are specified for each breeding cycle, such as number of parents (15), number of crosses per parent (14), and number of progeny from each parent (30). The founding population (cycle 0) therefore matches our progenitor population and serves as the backbone for 10 cycles of selection. The 10 cycles of selection are selected using truncation selection for 15 genotypes with the lowest phenotypic observation (objective to decrease the photoperiod causing flowering). The simulation is replicated 100 times and the mean (per cycle) phenotypic value and variance, along with their respective standard errors, are output.

Following the simulation of alternative genetic architectures, oligogenic control for flowering was selected to simulate alternative selection schemes for improvement of gain and maintenance of genetic variation through testing of genomic selection (GS), phenotypic selection with pedigree-based optimum contribution selection (Pedigree-OCS), and genomic selection with marker-based optimum contribution selection (GS-OCS). GS is employed through GBLUP by using a marker-based genomic relationship matrix to predict genomic estimated breeding values (GEBVs). Therefore, selections are based on GEBVs and not phenotypic performance, where phenotypes are estimated for selected individuals to serve as training for the following cycle of selection. The Pedigree-OCS framework is selection on the phenotype while allocating selections dependent on their predicted contribution to inbreeding in the following generation, minimizing pedigree-based kinship among selected individuals. The GS-OCS framework combines the methods of GS and Pedigree-OCS, where selection criteria is GEBV and predicted contribution to inbreeding of selections is minimized by marker-based identity-by-descent. For simplicity, our selection schemes aim to improve gain and/or maintain variation: (1) GS alters the selection criteria from phenotype to genotype, increasing selection accuracy; (2) Pedigree-OCS optimizes the selection criteria of phenotype to minimize inbreeding, increasing additive genetic variance; and (3) GS-OCS optimizes the selection criteria of genotype to minimize inbreeding, combining the benefits of the preceding schemes.

3.3 Results

Controlled environment trials to compare genotypes within and between generations (cycles of selection) exhibit differences in multiple domestication syndrome and stevia specific traits (Supplemental Figure 5). Selection has improved trait means in each breeding cycle, both observationally and statistically: leaf width (increase 5mm), leaf length (increase 5mm), internode spacing (decrease 0.5mm), photoperiod causing flowering (decrease 0.5hr), and steviol glycosides content (increase 3.5%); while other traits show expanded variation: branch count (range increase 5 branches) and plant height (range increase 13cm) (Supplemental Figure 5). Expanded variation is not only important to maintain gain, but since stevia is

clonally propagated, outliers can also be selected for production. The mean gain per cycle was 3.8% improvement for domestication traits and 5.1% for stevia specific traits. Correlations between TOIs range from highly positively correlated (TSG and RebA content: 0.9) to negatively correlated (Internode Spacing and Biomass Index: -0.4) (Supplemental Figure 6). Furthermore, the negative correlation between the photoperiod causing flowering and TSG/RebA contents ($r = -0.32, p < 0.001$) confirms the glycoside consumption during the ontogenetic change from vegetative to flowering (Bondarev et al., 2003).

3.3.1 Best Linear Unbiased Estimation (BLUE) of Generation Performance

A linear mixed-model (LMM) analysis using generation (cycle of selection) as fixed to estimate generation level performance of domestication syndrome and stevia specific traits of interest (TOIs) was used to explore selection efficiency. Domestication syndrome traits (dsTOIs) - flowering, plant height, branching, internode, leaf size – are estimated to improve over the 3 cycles of selection with mixed consistency (Table 5). For example, TOIs photoperiod causing flowering and internode spacing are found to have consistent improvement over each cycle of selection, a non-surprising result considering the correlation between them ($r = 0.3, p < 0.001$). However, other dsTOIs plant height, branching, leaf size is estimated to fluctuate over each generation, even though each respective dsTOI BLUE after three cycles of selection is improved over the progenitors (Table 5). This non-linear improvement, typically found at generation two, is a likely result of observational evaluation for these traits during the first and second cycles of selection as well as varying selection intensities. Stevia specific TOIs (ssTOIs) are estimated as improving over the progenitors, again with mixed consistency as observed in dsTOIs. For example, Rebaudioside A content has improved by nearly 3% after 3 cycles of selection with a slight decrease in the third generation while Total Steviol Glycosides content BLUE is improved by nearly 4% with gradual improvement over each cycle of selection. Estimated narrow-sense heritability for the TOIs are variable (Table 5). The largest heritability belongs to the dsTOI photoperiod causing flowering ($h^2 = 0.55$) with the lowest to dsTOI internode spacing ($h^2 = 0.13$). Lower than expected heritabilities are estimated for alternative dsTOIs, highlighting the imperfect selection in these traits and a by-product of error-prone observational evaluation.

Table 5: Best linear unbiased estimates of generation performance across domestication syndrome and stevia specific traits including variance components of random effects and estimated trait heritability.

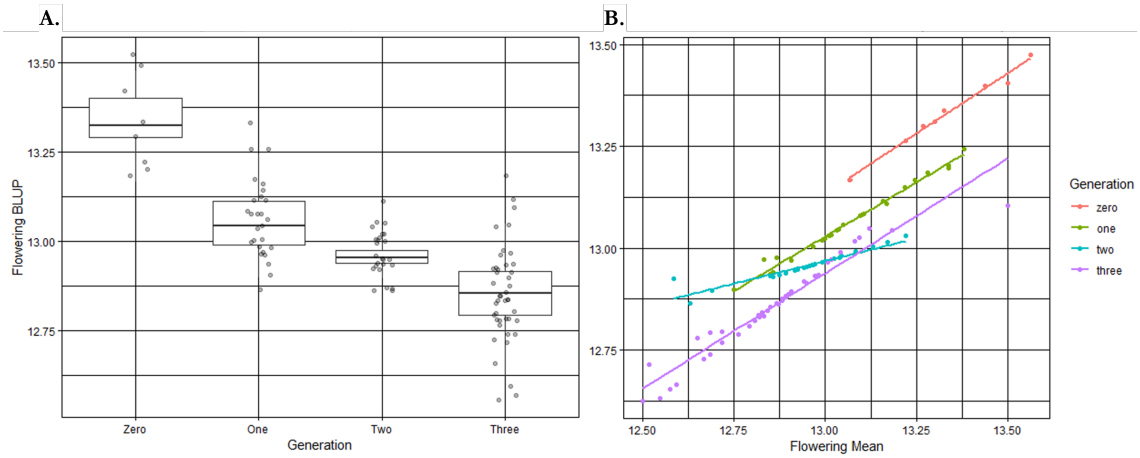
		Domestication Syndrome TOI						Stevia Specific TOI	
		Flowering (hr)	Plant Height (cm)	Branching (#)	Internode (mm)	Leaf Width (mm)	Leaf Length (mm)	Rebaudioside A (%)	Total Steviol Glycoside (%)
Generation BLUEs	Progenitor	13.32 (0.06)	17.82 (0.75)	8.73 (1.21)	2.16 (0.16)	23.40 (2.16)	54.63 (3.32)	10.91 (0.47)	17.32 (0.57)
	One	13.05 (0.05)	18.89 (0.74)	11.52 (0.94)	1.88 (0.16)	26.10 (1.58)	59.01 (2.74)	13.34 (0.33)	20.41 (0.39)
	Two	12.95 (0.06)	18.49 (0.77)	10.41 (1.05)	1.78 (0.17)	24.78 (1.84)	56.37 (3.14)	13.81 (0.43)	20.84 (0.51)
	Three	12.86 (0.06)	18.56 (0.74)	9.88 (1.05)	1.77 (0.17)	27.71 (1.85)	59.75 (3.14)	13.69 (0.44)	21.03 (0.53)
	Heritability (h2)	0.55 (0.15)	0.29 (0.06)	0.28 (0.07)	0.13 (0.05)	0.42 (0.06)	0.33 (0.06)	N/A*	N/A*
Var-Comps	Variance	0.01 (0.004)	0.65 (0.55)	4.39 (1.30)	0.06 (0.03)	18.96 (4.23)	47.76 (11.90)	1.84 (0.07)	2.65 (0.10)
	Covariance	0.04 (0.003)	11.01 (0.93)	13.54 (1.18)	0.45 (0.04)	28.47 (2.50)	97.89 (8.56)	N/A*	N/A*

*high-performance liquid chromatography performed on bulked sample across blocks, eliminating potential variation used for these components

3.3.2 Best Linear Unbiased Prediction (BLUP) of Variety Performance

Following our estimation of generation performance, another linear mixed-model using random predictors kinship informed variety and generation with heterogeneous variance was used to predict variety performance in dsTOIs. Variety BLUPs, informed through kinship, identifies 71 varieties with predicted photoperiod causing flowering at less than 13 hours of daylength with 7 being less than 12 hours and 45 minutes (Figure 11). However, using the mean (ignoring kinship information) observed photoperiod causing flowering, only 60 varieties are identified at less than 13 hours of daylength with 15 being less than 12 hours and 45 minutes. Therefore, the accuracy of the mean is less than that of BLUP, especially if we consider the best 10 genotypes of each method, 3 of which are identified by BLUP and not by mean as the selection unit. However, when leave-one-out cross-validation is used in training our model, predictive ability for photoperiod causing flowering drops to 0.20, highlighting the complexity of the trait and the underperformance of pedigree methodology.

Figure 11: (A) Boxplot of mixed-model best linear unbiased predictions of photoperiod causing flowering for genotypes grouped by generation and (B) their correlation to mean observed.



Variety BLUPs identifies 25 varieties with predicted number of at least 11 branches. However, using the mean observed number of branches identifies 33 varieties with number of at least 11 branches. When selecting using the BLUP branching, 2 of the top 10 varieties are different than selecting using the mean (Table 6). Variety BLUPs identifies 10 varieties for predicted plant height greater than 19cm. However, using the mean observed plant height identifies 49 varieties with greater than 19cm. Furthermore, only 6 of top 10 BLUP selected individuals for plant height are found in mean selected. Variety BLUPs identifies 13 varieties with greater than 30mm leaf width. However, using the mean observed leaf width identifies 26 varieties with greater than 30mm leaf width. This is the first dsTOI that does not have a large discrepancy between BLUP and mean selected, with only 1 variety of the BLUP selected individuals for leaf width (>30mm) being absent from the top 20 mean selected (Table 6). Variety BLUPs identifies 10 varieties with greater than 65mm leaf length. However, using the mean observed leaf width identifies 22 varieties with greater than 65mm leaf width. Similarly, to leaf width in the top 20, only 1 BLUP selected individual for leaf length is missing from

the top 20 mean (Table 6). Internode spacing variety BLUPs only identifies 4 varieties with less than 1.5mm internode while mean observed finds 19 varieties. Only 3 of the top 20 varieties from BLUP are missing from mean observed.

Table 6: Generational breakdown of truncated selection of top 20 best linear unbiased predicted and mean performing varieties for domestication syndrome traits of interest.

TOI	Generation	BLUP	Mean
Flowering	Progenitor	0	0
	One	2	1
	Two	1	3
	Three	17	16
Plant Height	Progenitor	0	0
	One	9	8
	Two	6	3
	Three	5	9
Branching	Progenitor	1	0
	One	7	9
	Two	8	7
	Three	4	4
Internode	Progenitor	2	2
	One	9	7
	Two	1	1
	Three	8	10
Leaf Width	Progenitor	0	0
	One	2	2
	Two	3	3
	Three	15	15
Leaf Length	Progenitor	0	1
	One	3	3
	Two	5	5
	Three	12	11

3.3.3 Stochastic simulation for future projection and genetic complexity

Stochastic simulation was used to project potential gains in flowering time trait. More variable phenotypic gain is estimated than phenotypic variance, comparing genetic architectures (Figure 12A/B). Over the simulated 10 cycles of selection, the oligogenic trait exhibits the greatest overall mean gain (almost 1 hour decrease) as well as the second worst variance loss (about 0.01). Since there is a single environment used in evaluation (reality and simulation), polygenic trait gain is inefficient and sports only a half-hour gain over 10 cycles. When BLUEs are placed over mean phenotypic value simulated forecast, we observe nearly perfect continuity, despite generations 1-3 BLUEs not being input anywhere in the simulation (Figure 12A/B). When BLUE variances are placed over mean phenotypic variance simulated forecast, we find the points to be within the lower bounds of standard error for the simulation replications, a likely outcome due to progressing inbreeding in our population (Figure 12B). Considering these results, we propose the genetic architecture of photoperiod causing flowering to be oligogenic, similar to reported genetic complexity in flowering traits of alternative species (Jungers et al., 2023).

Once genetic architecture is chosen, alternative selection schemes were explored by altering selection criteria for optimization (Figure 12C/D). Phenotypic selection simulated gain over 10 cycles of selection is 6.3%, where the mean phenotype at 12.48 hours of daylight causing flowering. GS simulates reduced gain over 10 cycles of selection by almost 50% and reduces genetic variation more than phenotypic selection, a likely result given the simulation format of a single environment trial which typically limits trait heritability. Pedigree-OCS

improves gain over GS and maintains half-way as much variation in cycle 10 as GS and phenotypic selection with 4.4% improvement to mean of 12.73 hours. GS-OCS does not improve gain over Pedigree-OCS but does over GS as well as maintaining as much variation in cycle 10 as phenotypic selection. When we increase environments of phenotyping (Figure 12E/F), gain is improved in GS by 1.2% and Pedigree-OCS by 0.4% but is unimproved in GS-OCS and declines in phenotypic selection by 1.0%. This increase in gain is less than reported in intermediate wheatgrass (Zhang et al., 2016). However, genetic variation is severely lost with the improved selection accuracy derived from multi-environment phenotyping, likely requiring genetic introgression shortly after 10 cycles to increase trait variability for future population improvement.

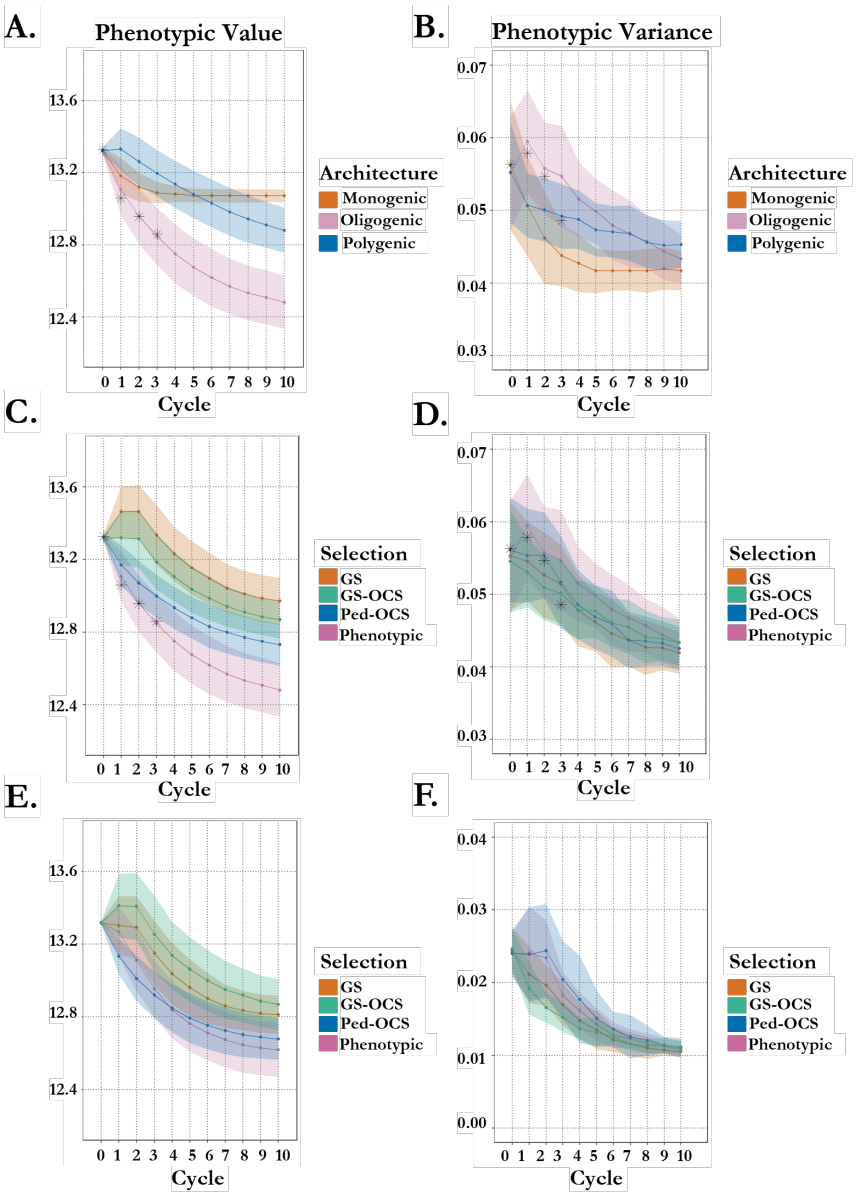
3.4 Discussion

3.4.1 Implications for selection in a semi-domesticate

Domestication is population improvement towards agroecological adaptability and agronomic performance (Harlan et al., 1973, 1975). Breeding cycle components can be manipulated to generate sustainable gain (Cobb et al., 2019). Our case study into selection in stevia outlines progress of domestication TOIs is possible through phenotypic recurrent selection. Moreover, our approach to augment selection schemes through stochastic simulation identifies methods and technology to increase gain and maintain more genetic variation during this domestication phase. Implementation of GS can improve gain per cycle, and with potentially 3 cycles per year by also leveraging speed breeding technologies makes progress towards domesticate form rapid (Jannink, 2010, Watson et al., 2018). Large selection intensity has the potential to induce rapid bottlenecking in the population under selection, especially when integrating technologies that make that selection more accurate (GS). This loss of variation could be problematic with changing climatic conditions and continual evolution of pests and pathogens. However, historical bottleneck were potentially not as bad as we previously thought, albeit under phenotypic selection (Allaby et al., 2019).

Artificial selection drives breeding population improvement, just as natural selection of heritable mutational variants drives evolution and speciation (Lande, 1976). But, we are making the decisions with the goal of domestication being development of a population with agroecosystem adaptation. Selection methods and technologies to improve genetic gain (Figure 1) should be leveraged. However, careful consideration of genetic variation is key, to preserve future gain of more complex syndrome traits. Genetic variation can be lost at lower rates by implementing optimum contribution selection and increased by genetic rescue, two approaches that ought to be used during domestication (Allier et al., 2020, Bančič et al., 2021, Bertorelle et al., 2022). Historical domestication reports of multiple domestication events for some crops (Kovach et al., 2007) where populations were selected separately and subsequently combined, a potential technique that should be considered for boosting genetic variation in neo-domestication. This technique paired with marker-assisted backcrossing (simple traits) or genomic selection (complex traits) could alleviate some of the expected outcrossing depression during population combination (Lynch, 1991, Allier et al., 2020). Neo-domestication is ripe for empirical testing because, as breeders, we have a constantly

Figure 12: 100 stochastic simulations of photoperiod causing flowering in stevia under varying genetic architecture, selection criteria, and environments of evaluation. The plotted points and lines represent the mean of replications per cycle with the shaded regions representing the standard deviation of replications per cycle. (A/B) Varying the genetic architecture of photoperiod causing flowering using phenotypic recurrent selection and a single environment. (C/D) Varying the selection criteria for photoperiod causing flowering in stevia under oligogenic control and a single environment. (E/F) Varying the selection criteria and increasing the number of environments to four during evaluation of photoperiod causing flowering in stevia under oligogenic control. (A/C/E) represent the phenotypic gain, measured as phenotypic value and (B/D/F) represent the phenotypic variance. Selection criteria include genomic selection (GS), marker-based optimum contribution selection (GS-OCS), pedigree-based optimum contribution selection (Ped-OCS), and phenotypic recurrent selection (Phenotypic).



evolving understanding and knowledge around the historical and future stages of breeding, as mentioned by (Wallace et al., 2018): domestication, quantitative genetics theory, molecular markers, functional variant combination, and custom genetic design. The decision of which species to use should take into consideration those semi-domesticated by using a pipeline strategy which considers potential market capture (DeHaan et al., 2016, Fernie and Yan, 2019). These considerations make stevia an excellent study species because of its novel market application and semi-domesticated form (Matsui et al., 1996, Clemente et al., 2021).

3.4.2 Potential for use of Selection Index for traits

Neo-domestication is plant breeding utilizing modern techniques and methodologies (e.g. 5 steps listed by (Wallace et al., 2018)) and plant breeding is based on selection of traits that are desirable for humans (Bernardo, 2014). Selection is often implemented on individual traits, but the improvement of multiple traits through an index is the most effective method for improving those traits simultaneously (Hazel and Lush, 1942). Selection indices have been a part of breeding for decades (Lush, 1935, Smith, 1936, Hazel, 1943, Lush et al., 1949, Henderson, 1975). Application of a selection index to the species identified domestication syndrome should be considered to improve all relevant traits simultaneously, an area for future investigation.

For example, we can take our population improvement program in stevia to understand the discrepancy between selection for one trait at a time versus selection for a total score (e.g. selection index). Progress for any one trait by the total score method is only $1/\sqrt{n}$ times (n=number of traits) as much as if selection were directed to that trait alone (Hazel and Lush, 1942). To begin this comparison, we can calculate expected genetic gain or phenotypic gain for 2 traits (daylength causing flowering and leaf length) separately, where gain is equal to the product of the intensity of selection, accuracy of selection or narrow-sense heritability, and additive variation or phenotypic variation (Lynch and Walsh 1998, Rutkoski 2019). First, we will calculate selection intensity (i) like methods proposed by (Falconer and Mackay, 1996) by taking a linear approximation of i. Our methods thus far have been producing roughly 5,000 new progeny each generation and selecting 30 individuals for breeding (selection proportion= 0.6%). Therefore, the linear approximation of our selection intensity finds i=2.834. Second, we will take our linear mixed-model narrow-sense heritabilities (flowering=0.55 and leaf length=0.33; Table 1). Third, we need the phenotypic variance (flowering=0.048 and leaf length=170.50) to get standard deviation of phenotype (flowering=0.219 and leaf length=13.06). Response to selection is therefore estimated on single trait selection in the fourth generation to be 0.34 in flowering or 12.21 in leaf length. However, applying the reduction in gain when selecting for multiple traits as being adjusted by $1/\sqrt{n}$ (n=2), gain on an index for flowering and leaf length would be 0.24 and 8.63 (Hazel and Lush, 1942).

3.5 Conclusion

Linear mixed-model BLUPs have identified 7 lines being less than 12 hours and 45 minutes daylength causing flowering, 12 lines with greater than 30 mm leaf width, 13 lines with

greater than 65 mm leaf length, 5 lines greater than 12 branches, and 10 lines with greater than 19 cm in height. There are 11 lines that are found for more than one dsTOI out of the 47 lines at the above trait-specific culling levels. These 11 lines trace their lineage back to 5 of the 8 progenitors (1 from A, 1 from C, 2 from D, 3 from E, 3 from F, and 1 from H). As we move the breeding program forward, consideration should be given to alternative selection criteria: (1) Pedigree-OCS to maintain more genetic variation in our breeding populations; (2) implementation of GS or GS-OCS, requiring genotyping-by-sequencing and an increase in trial environments improve predictive accuracy; (3) selection using an index, requiring trait-specific economic valuations. These shifts are in no specific order and rather speak to the different potential approaches for improving gain and/or maintaining variation

Chapter 4: Optimizing cost efficiency under neo-domestication and wide hybridization breeding schemes

4.1 Introduction

Domestication can broadly be defined as coevolution through selection to improve the adaptation of plants for human cultivation and preferences. Historically this has been done to accumulate yield and productivity, harmonize crop management practices, promote ease of harvest, and improve palatability. Crop domestication imposes several microevolutionary forces on the plant genome in comparison to their wild ancestors: (1) selectively neutral forces (genetic drift and gene flow) are expected to have genome-wide effects with genetic drift decreasing genetic diversity and gene flow maintaining or increasing genetic diversity; and (2) selection leads to differential loss of genetic diversity in targeted genomic regions, creating a molecular signature of selection (Olsen and Wendel, 2013). Furthermore, domestication alters selection pressures so that wild favored traits become neutral or disfavored, results in purifying selection following bottleneck, and reduces effective population size and effective recombination rates through inbreeding (Wallace et al., 2018). Thousands of plant species spanning 160 taxonomic families have undergone some extent of domestication (Meyer et al., 2012, Meyer and Purugganan, 2013). The total number of cultivated consumable species is ∼7,000 (Khoshbakht and Hammer, 2008), exemplifying the massive botanical resource for improving agricultural sustainability.

Historical domestication is postulated to have started ∼12,000 years ago with intensive agricultural production beginning some millennia later (Meyer and Purugganan, 2013, Fuller et al., 2014, Purugganan, 2019). The 'domestication syndrome' refers to the set of phenotypes for traits which improve adaptability to the human agroecosystem compared to wild type (Harlan et al., 1973). The domestication syndrome varies by crop-type, where seed-propagated annuals typically obtain a different syndrome compared to fruit trees, vines, and tubers (Zohary and Spiegel-Roy, 1975, Zohary et al., 2000, Gaut et al., 2015). Syndrome traits not only vary by crop-type, but also by genetic architecture, ranging from monogenic to polygenic control (Jungers et al., 2023). The target traits for neo-domestication are the same as under historical domestication but with an improved understanding of the architecture and effects of selection. Typically, these traits are of monogenic (transposable elements such as tb1 for branching in maize or Gret1 for anthocyanin synthesis in grapes), oligogenic (seed shattering, seed size, dormancy traits), and polygenic (flowering, kernel composition, leaf morphology, resistances) control (Jungers et al., 2023). Archaeological genetic evidence suggests a multi-stage process to domestication through the increase of favorable alleles, creation of cultivated populations, and breeding (Fernie and Yan, 2019). Despite this process being outlined as three stages, the progress was extremely slow owing to the presence of genetic migration from the wild, inaccurate and often unconscious selection, increases in genetic

load through bottleneck, and more. However, modern breeding technologies and methodologies have the potential to speed the progress through these stages enacted under historical domestication (Watson et al., 2018, Zhang et al., 2023).

Climate extremes are decreasing food security, exacerbated by the increasingly globalized food system, where localized production using crop-types amenable to locales enriches the sustainability of regional food systems. Although adaptability in major food crops has shown some success, historically, production and consumption of species is regional (e.g. wheat in Europe, rice in Asia, maize in the Americas). Even though global trends of consumption prefer access to all these major grain crops, regional types of plant species can provide sustainability to their locale (tef, sorghum, amaranth, mung bean, etc.), especially in the current variable climate. Broad adaptation strategies towards a changing climate and production region shifts include: (1) sourcing crop populations (e.g. landraces, varieties) from different global geographic regions matching future projected climate, (2) assessing crop wild relatives for naturally evolved adaptations, (3) defining replacement crops to be cultivated, (4) defining different agroecosystems for existing crops, (5) substantially changing agronomic practices such as row spacing, irrigation and planting date, and (6) abandoning current production locations with human population moving to areas amenable to current practices/cultivars (Burke et al., 2009, Ramirez-Villegas and Khoury, 2013, Pironon et al., 2019, Sloat et al., 2020). Our study goal is to develop the breeding schemes relevant towards population development through recurrent selection to include wild and crop wild relative species to improve the sustainability of the local and regional food system.

A proven approach to increasing genetic diversity and adaptability in crop species is through the utilization of wild relatives for crop improvement (Jansky et al., 2013, Dempewolf et al., 2017, Mehrabi et al., 2019). Utilizing wild relative species, as well as landrace and heirloom varieties, provides a mechanism to alleviate the abiotic stresses expected with climatic shifts through evolved traits including tolerance to salinity, drought, and temperature extremes (Bailey-Serres et al., 2019, Ramankutty et al., 2018, Flint-Garcia et al., 2023). Historic climatic events and shifts have placed pressure on crop cultivars by creating novel abiotic and biotic stresses (Lesk et al., 2016) for which natural resistance variation in wild populations exist. This variation can be leveraged to improve existing crops or to develop novel breeding populations. Understanding the progress of these breeding populations can be observed and predicted through phenotypic and genetic values.

The genetic gain equation is the product of additive genetic variation within a population, selection intensity, and selection accuracy divided by cycle length used to predict the response to selection (Lynch et al., 1998, Cobb et al., 2019). The breeder's equation is an important development in quantitative genetics because it provides a framework for predicting the response to selection within a population (Lynch et al., 1998). Therefore, a breeder's game theory can be developed and optimized prior to specific action by predicting a population's response (e.g. genetic gain and variance) to artificial selection methods. Traits of interest under modern neo-domestication efforts vary from monogenic through polygenic, often with the ideotype possessing no variation for simple traits (non-shattering, branching, dormancy) to fix within the population whilst also possessing enough variation for more complex traits (flowering, leaf morphology, resistances) to maintain adaptive potential in the population. It is often these simple traits that inhibit human mediated cultivation while the complex traits promote expansion of cultivated area. Therefore, fixation of simple traits and preservation of

additive genetic variation of complex traits should be the goal of neo-domestication programs.

Stochastic simulation offers a unique framework to identify breeding cycle components which optimize the target (gain, variance, cost, etc.) (Bernardo, 2020). Varying parameters across the breeding cycle produces variable results for the given target which can be used to strategically design a breeding scheme. Breeding schemes are the combination of crossing, evaluation, and selection component parameters used to maximize genetic gain in a breeding population per dollar invested (Covarrubias-Pazaran et al., 2022). Crossing parameters include variables such as number of parents, number of crosses, number of progeny, type of cross, and mate allocation. Evaluation parameters include variables such as number of locations, replications, number of checks, experimental design, plot sizes, and subsamples. Selection parameters include percentage of selected individuals (selection intensity), selection method (culling, tandem, index), and the selection unit (phenotype, GEBV, etc.). Genetic gain can be improved through specific focus on components of the breeders' equation. Selection intensity can be increased by increasing the selection candidates while holding those selected constant and selection accuracy can be increased primarily through increasing heritability of the trait of interest. The number of cycles per year can be increased through predictive methodologies (e.g. genomic selection) and novel strategies of increasing plant development with longer daylength (e.g. speed-breeding). Additive genetic variation is important to maintain gain for quantitative traits when using predictive methodologies (Jannink, 2010), which could be increased or maintained through mate allocation. However, additive genetic variation typically attenuates as selection accuracy increases, directly impacting potential gain (Olsen and Wendel, 2013, Wartha and Lorenz, 2021).

There are numerous examples of stochastic simulations used for breeding scheme development in plants (Wang et al., 2003, Gaynor et al., 2017, Gorjanc et al., 2018, Allier et al., 2020, Bančič et al., 2021) and animals (Wall et al., 2010, Villalba et al., 2019). Wang et al. compares the pedigree/bulk selection method and the selected bulk selection method to understand genetic gain differences accounting for epistasis, pleiotropy, and GxE interaction through simulation (Wang et al., 2003). Gaynor et al. uses simulation to investigate a two-part strategy of product development, focusing on developing and screening inbred lines, and population improvement, focusing on increasing the frequency of favorable alleles through rapid recurrent genomic selection (Gaynor et al., 2017). Gorjanc et al. expands the two-part strategy using optimal contribution selection to investigate how this technique of mate allocation reduces the rapid loss of genetic variation during recurrent selection (Gorjanc et al., 2018). Allier et al. used simulation to investigate methods of pre-breeding and introduction of genetic diversity during population improvement as well as applying specific mate allocation techniques (Allier et al., 2020). Bančič et al. even expands the single crop simulations to investigate intercrop breeding by using an alteration of specific and general combining abilities (Bančič et al., 2021). Simulations have been used to integrate emerging technologies such as phenomics (Peixoto et al., 2023) and to optimize resource allocation (Lorenz, 2013, Ben-Sadoun et al., 2020, Jannink et al., 2023). However, this scheme development research typically uses major crops as their model and only some examples integrate costs which are limited in scope, focusing on hybrid development of major crops (Bernardo and Yu, 2007, Lorenz, 2013, Jannink et al., 2023, Peixoto et al., 2023). Moreover, these cost-integrated simulations typically only provide a single cost per methodology or technology, instead of providing a range of costs that realistically would be undertaken depending on

region, resource availability, organizational structure, budget, and crop-type.

Therefore, the goal of this study is threefold. First, we design in-silico populations to match potential germplasm acquisition scenarios from wild, orphan (semi-domestic), or landrace populations, each formed through varying effective population sizes and population mating sizes over 40 cycles of varied selection intensity and genetic drift (i.e. burn-in phase). Second, the variable populations developed in burn-in phase are incorporated into contemporary breeding, altering breeding cycle parameters to document phenotypic gain on two different architecturally controlled traits under variable breeding schemes (i.e. population development phase). Third, we apply a range of potential costs for breeding scheme methodologies and technologies applied in each parameter combination to understand the return on investment for a given population type and scheme.

4.2 Materials and Methods

Stochastic simulations were used to compare the return on investment (ROI) of different population development pipelines using in-silico wild and semi-domesticated types of species. We tested different population complexities, breeding cycle parameters, and costs to formulaically identify parameters which optimize phenotypic gain given different populations under variable resource availability. Phenotypic gain is used instead of genetic gain because incipient domestication studies often lack genetic resources, serving to make findings applicable to researchers in the field and valuable for comparison to on-going neo-domestication programs. Each simulated scheme is compared using the mean of 3 replicates, with each consisting of: (1) a burn-in phase of 40 cycles to allow for genetic drift, mutation, and variable selection for population formation of which germplasm can be collected; and (2) a breeding phase of 40 cycles of selection from differential populations for population development breeding schemes through leveraging alternative breeding cycle parameter combinations. After all schemes are simulated, a range of costs associated with the methods applied are designated and used to calculate the ROI.

4.2.1 Simulation of founding populations and trait genetic values

The genome was simulated as consisting of 10 chromosomes and varying effective population sizes (e.g. Ne = 25, 50, 100). The recombination and mutation rates were $1.25x10^{-8}$ and $2.5x10^{-8}$ per base pair, respectively (Hickey et al., 2014). Founder genotypes were generated using the Markovian Coalescent Simulator housed in the AlphaSimR package (Chen et al., 2009, Gaynor et al., 2021). The genome was simulated with 10,000 single nucleotide polymorphisms (SNP) with each population consisting of two simulated traits: (1) simple oligogenic (z_1) with 8 quantitative trait loci (QTL); and (2) complex oligogenic (z_2) with 400 QTL. Genetic values were simulated as two independent traits with mean and variance at zero and one, respectively. The narrow-sense heritability (h2) varied for each trait independently through different pairs (e.g. $h^2 = 0.7, 0.3; 0.3, 0.7; 0.3, 0.3; \&0.5, 0.5$). Both traits were simulated using additive effects for each QTL and were sampled from a normal distribution. The phenotypic value of each simulated genotype was calculated by adding the random sampled error for the genetic values of each trait and sampled from a normal distribution with

mean zero and residual variance based on each trait's h^2.

4.2.2 Differential population formation - burn-in phase

The burn-in phase used different selection depending on the population type to serve as the baseline of potential germplasm acquisition scenarios. Each population type was formed through 40 cycles of burn-in under varying selection: (1) the wild population formation used random selection to allow for genetic drift; (2) the orphan population formation used lenient truncation selection to allow for slow directional selection and genetic drift; and (3) the landrace population formation used strict truncation selection to allow for quick directional selection and genetic drift. Truncation selection used in (2) and (3) is based on a weighted index to account for dual trait selection where the weight for z_1 and z_2 is 0.5. The number of parents, crosses, and progeny per cross varies within each population type (Table 7). This factorial design will provide insight into the size of a population in which germplasm is acquired and its role in phenotypic gain along the gradient of Ne. The resulting population from each parameter iteration is then passed to the population development pipeline by selecting individuals from cycle 40.

Table 7: Burn-in phase crossing parameters

# of Parents	# of Crosses	# of Progeny/Cross	Total Progeny
20	190	2	380
		10	1900
		20	3800
		40	7600
30	435	2	870
		10	4350
		20	8700
		40	17400
40	780	2	1560
		10	7800
		20	15600
		40	31200
50	1225	2	2450
		10	12250
		20	24500
		40	49000

4.2.3 Population development - breeding phase

The breeding phase used different breeding cycle parameters, including but not limited to population size, crosses made, progeny evaluated, and selection criteria. Each potential combination of parameters was tested on each population formed during the burn-in phase for a total of 5,184 unique combinations each with 3 replications. Population development breeding phase begins with selection of individuals based on weighted index of z_1 and z_2

($si_w = 0.25, 0.75; 0.5, 0.5; 0.75, 0.25$). We then apply phenotypic recurrent selection for the first two cycles (cycle 41 and 42) to serve the role of data collection and germplasm knowledge that is required when beginning a breeding program because burn-in phase data and germplasm is considered unknown in our framework. The entire breeding phase consists of 40 cycles of selection, including cycles 41 and 42, where the latter 38 cycles apply phenotypic recurrent selection (PRS), genomic selection (GS), or maximum avoidance (MxAv). The selection criterion is still the weighted index of z_1 and z_2 referenced above.

The first selection criteria scenario simulated across all parameter combinations is PRS, serving as the baseline for our study. We then apply GS through genomic relationship, retraining every other cycle, to understand the impact of gain for the two traits by selection on the breeding value and decreasing the length of time per cycle (2 cycles per year) using the R package 'sommer'. GS is well understood to rapidly decrease genetic variance, we apply MxAv based on pedigree relationships to understand the impact of variance for the two traits using the R package 'optiSel'. Maximum avoidance (MxAv) selection is essentially using optimum contributions to select individuals based on minimizing inbreeding, but ignoring proportion of contribution and instead using circular mating for equal representation and static population size (Kimura and Crow, 1963). The strategic outlook for these methods has implications towards increasing the rate of gain through GS's ability to increase selection accuracy and decrease time per cycle or decreasing the loss of variance through MxAv's ability to minimize mean kinship of progeny by reducing the inbreeding coefficient. Similarly, to the burn-in phase, the number of parents, crosses, and progeny per cross varies within each population type (Table 8).

Table 8: Population development phase crossing parameters

# of Parents	# of Crosses	# of Progeny/Cross	Total Progeny
10	45	1	45
		5	225
		10	450
		20	900
15	105	1	105
		5	525
		10	1050
		20	2100
20	190	1	190
		5	950
		10	1900
		20	3800
25	300	1	300
		5	1500
		10	3000
		20	6000

4.2.4 Applying cost to the parameter combinations and return on investment formulation

The costs associated with specific breeding schemes is a crucial parameter towards strategically optimizing the breeding target. Despite its importance, the parameter is often restricted to a single value or left out of consideration altogether. Here, we use a range of costs (Low, Medium, and High) for a range of crop-types (Field, Horticultural, and Forestry) to develop an understanding of the impact on breeding scheme strategies (Table 9). The costs are based on incipient domestication and wide hybridization programs housed at the Hawaii Agriculture Research Center and strictly restricted to phenotyping and/or genotyping costs (Table 3), as the Medium cost per plot and/or genotyping sample for each crop-type. The costs per plot and/or genotyping sample is then multiplied by the entire cycle population size to gather a yearly phenotyping/genotyping cost and then summed for the entire cost of a specific scheme across 40 cycles of selection. The breeding targets (gain and variance) are compared by computing their change over the 40 cycles of selection and scaling towards our baseline parameter combinations within each scheme: phenotypic recurrent selection, $h^2_{z1} = 0.7$ and $h^2_{z2} = 0.3$, equal selection index weighting (0.50) per trait, and orphan crop population. Scaling is done by dividing each parameter combination within every scheme by the baseline to create a unit-less approach to understanding the change in gain and variance. This technique is applied for the range of costs as well, using the same baseline combination. Return on investment (ROI) is then viewed as a unit change in gain or variance for each trait per unit cost.

Table 9: Costs per plot and/or genotyping sample. Medium level costs for each crop-type of field, horticultural, and forestry are based on the phenotyping costs for stevia, cacao, and koa, respectively.

Cycles	Selection Type	Field Crop			Horticultural Crop			Forestry Crop		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
2	GA	5	8	10	10	15	20	20	25	30
38	PRS	10	15	20	20	30	40	40	50	60
38	GS	10	17	25	15	25	35	25	35	45
38	MxAv	10	15	20	20	30	40	40	50	60

4.2.5 Mixed-model regression for optimizing ROI

Following the scaling of breeding targets and costs as well as ROI formulation, we fit a linear mixed model for best linear unbiased estimation of this unit change per unit cost for every stochastic simulation scheme using the R package "lme4". Our model follows the form of ROI as the response variable with fixed predictors including population type, selection criteria, heritability, selection index weighting, and the interactions of heritability with selection index weighting and selection criteria. Random predictors include the starting effective population size, the number of parents, and the number of progeny per cross.

4.2.6 Calculating the asymptote of both traits' phenotypic gain and variance

Although the unit change in gain or variance per unit cost is useful for comparison of breeding schemes for neo-domestication breeding programs, the goal of these programs should also be to fix simple traits with high phenotypic value within the population while maintain-

ing enough variation for complex traits with high phenotypic value to facilitate sustainable improvement. To parse this, we fit sigmoidal or logistic curves along every scheme to identify the breeding cycle asymptotes of phenotypic value (upper) and variance (lower) for z_1 and z_2. We apply through nonlinear least squares a sigmoidal curve by four-parameter logistic model along the cycles of selection for both targets of both traits for every scheme combination for 20,736 asymptote locations (5,184 schemes each with 4 targets). Starting points of upper and lower asymptotes were specified as the minimum or maximum value of target per scheme. Approximately 75% of the asymptotes were identified, with the remaining being unidentified, where we moved to applying a linear model for predicting the unidentified asymptote using scheme parameters and the sigmoidal midpoint ($r = 0.68; p < 0.001$).

4.3 Results

Breeding schemes showed a wide range of phenotypic gain and variance depending on selection strategy and population type. For example, the largest phenotypic gain for both z_1 and z_2 during population improvement were under PRS for landrace and orphan populations but GS for wild populations (Table 10). The smallest phenotypic gain, sometimes a loss of phenotypic value, were typically under MxAv selection. The effect of selection on variance varied by population type and trait: (1) the largest increases in variance for trait z_1 were GS for landrace and orphan populations with MxAv increasing variance in the trait for wild populations; and (2) the variance for trait z_2 was increased with MxAv in orphan and wild populations with PRS increasing variance for the trait in landrace populations (Table 4). The largest decreases for variance for the z_1 were observed as PRS for landrace and GS for orphan and landrace populations while for z_2 variance was decreased by GS in wild, PRS in orphan, and MxAv in landrace.

The starting effective population size and population type has direct influence on the mean gain of both traits, z_2 mean phenotypic gain increases within each population type alongside increasing effective population size of germplasm acquired while z_1 mean phenotypic gain does not increase with effective population size but larger gains are observed when germplasm is acquired from landrace or orphan populations (Supplemental Figure 7). The narrow-sense heritability of each trait impacts mean phenotypic gain of z_2 while having minimal impact on gain of z_1 over the 40 cycles of selection during population development (Supplemental Figure 8). Selection and the starting effective population size play a role in increasing mean phenotypic gain during population development, where all selection methods similarly improve z_1 despite Ne (Supplemental Figure 9). However, z_2 has the largest gains under GS and PRS with large Ne, while MxAv shows smaller mean gain. The weighting of each trait shows a larger impact on gains of the complex z_2 trait with minor effects on the simple z_1 trait (Supplemental Figure 10).

4.3.1 Best linear unbiased estimates of unit by unit cost change across targets, schemes, and cost ranges

The best linear unbiased estimates (BLUEs) of unit change in the target varies by the cost of phenotyping, genotyping, and breeding schemes applied. The mean z_1 BLUEs of gain

Table 10: Largest increases and decreases of targets by population type. Unit change represents the trait & target value per cost scaled to the baseline scheme trait & target per cost value (phenotypic recurrent selection, $h^2_{z1} = 0.7$ and $h^2_{z2} = 0.3$, equal selection index weighting (0.50) per trait, and orphan crop population). Increase/decrease represents the largest increase/decrease in unit change per population type across all trait & targets.

Direction	Population	Ne	Nprogeny	Nparents	h2	Selection	si weight	Trait & Target	Unit Change
Increase	Landrace	25	10	15	0.3	PRS	0.75	z1 gain	1.73 (61.7%)
		25	1	10	0.3	GS	0.25	z1 variance	1.27 (58.9%)
		100	20	20	0.7	PRS	0.75	z2 gain	13.64 (112.9%)
		25	1	10	0.7	PRS	0.5	z2 variance	1.21 (46.4%)
	Orphan	25	20	15	0.3	PRS	0.5	z1 gain	3.74 (81.8%)
		100	1	10	0.3	GS	0.5	z1 variance	1.53 (68.1%)
		100	20	25	0.3	PRS	0.75	z2 gain	18.18 (251.9%)
		25	1	10	0.3	MxAv	0.75	z2 variance	1.09 (41.6%)
	Wild	50	5	25	0.3	GS	0.75	z1 gain	15.89 (7716.5%)
		50	1	10	0.3	MxAv	0.25	z1 variance	0.94 (35.7%)
		100	20	25	0.7	GS	0.75	z2 gain	22.93 (3626.0%)
		100	1	15	0.3	MxAv	0.5	z2 variance	0.77 (30.5%)
Decrease	Landrace	25	20	15	0.5	PRS	0.25	z1 gain	-1.19 (-39.0%)
		25	1	10	0.3	PRS	0.5	z1 variance	-1.45 (-56.3%)
		50	1	10	0.3	GS	0.25	z2 gain	-0.14 (-1.3%)
		50	1	10	0.3	MxAv	0.5	z2 variance	-1.80 (-58.4%)
	Orphan	25	10	15	0.3	MxAv	0.25	z1 gain	-1.61 (-42.4%)
		50	1	10	0.3	GS	0.75	z1 variance	-1.71 (-52.3%)
		25	1	10	0.3	MxAv	0.25	z2 gain	0.15 (5.8%)
		50	1	10	0.3	PRS	0.25	z2 variance	-2.13 (-67.4%)
	Wild	50	10	10	0.3	MxAv	0.25	z1 gain	-1.70 (-12235.5%)
		100	10	15	0.5	GS	0.5	z1 variance	-2.94 (-74.6%)
		100	10	10	0.3	MxAv	0.25	z2 gain	0.37 (116.4%)
		100	1	10	0.3	GS	0.75	z2 variance	-2.52 (-54.5%)

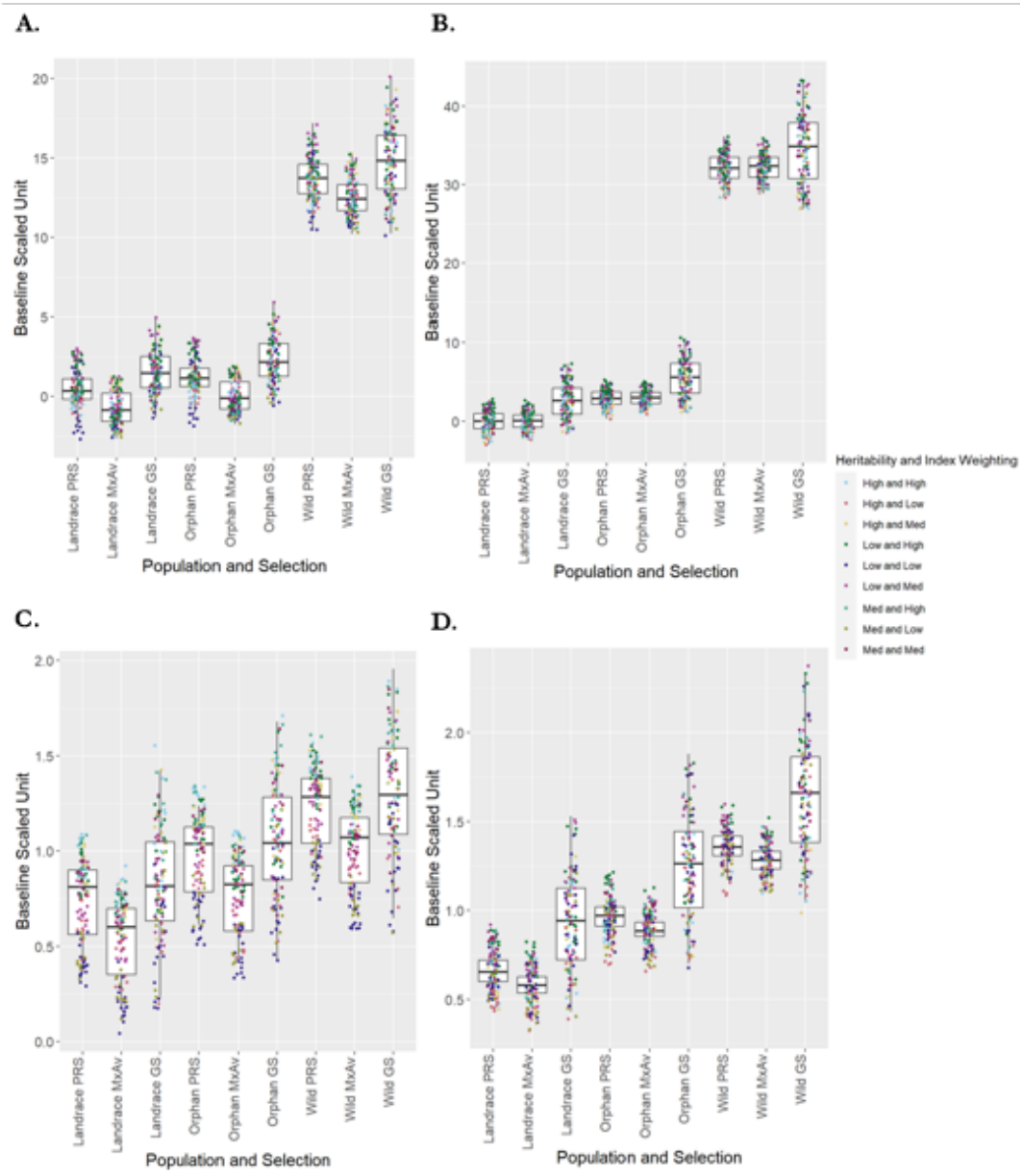
target finds GS with 0.3 heritability and 0.5 selection index weighting to be ∼3 times greater than baseline selection scheme in landrace populations, GS with 0.7 heritability and 0.25 selection index weighting to be ∼2.5 times greater in orphan populations, and GS with 0.7 heritability and 0.25 weighting to be ∼15 times greater in wild populations. The mean z_1 BLUEs of variance loss target finds PRS with 0.7 heritability and 0.25 weighting to be ∼1.5 times less than baseline in landrace and orphan and ∼30 times more in wild populations. The mean z_2 BLUEs of gain target finds GS with 0.7 heritability and 0.75 weighting to be 6% greater than baseline in landrace populations, 29% greater in orphan populations, and 53% greater in wild populations. The mean z_2 BLUEs of variance target maintenance finds MxAv with 0.7 heritability and 0.25 weighting to lose 50% less than baseline in landrace, 20% less variance loss in orphan, and ∼20% more variance loss in wild populations.

Figure 13: Linear mixed-model analysis of variance covariate significance heat-map by cropping system, cost level, trait target, and covariate. Identifying the level of significance by Type III ANOVA with Satterhwaite's method. $p < 0.001$ is filled with green, $p < 0.01$ is filled with yellow, $p < 0.1$ is filled with orange, and no significant effect is filled with red.

	Field											
	Low				Medium				High			
Covariate	Simple Gain	Simple Variance	Complex Gain	Complex Variance	Simple Gain	Simple Variance	Complex Gain	Complex Variance	Simple Gain	Simple Variance	Complex Gain	Complex Variance
Population	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Selection Criteria	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p<0.1	p<0.001	p<0.001	p>0.1	p<0.001	p<0.001	p<0.001
Heritability	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
SI Weighting	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
Heritability : Selection Criteria	p>0.1	p>0.1	p<0.01	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
Heritability : SI Weighting	p>0.1	p>0.1	p<0.001	p<0.1	p>0.1	p>0.1	p<0.001	p<0.1	p>0.1	p>0.1	p<0.001	p<0.1
	Horticulture											
Population	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Selection Criteria	p<0.1	p<0.001	p<0.001	p<0.001	p>0.1	p<0.001	p<0.001	p<0.001	p>0.1	p<0.001	p<0.001	p<0.001
Heritability	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
SI Weighting	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
Heritability : Selection Criteria	p>0.1	p>0.1	p>0.1	p<0.001	p>0.1	p>0.1	p>0.1	p<0.001	p>0.1	p>0.1	p<0.1	p<0.001
Heritability : SI Weighting	p>0.1	p>0.1	p<0.001	p<0.1	p>0.1	p>0.1	p<0.001	p<0.1	p>0.1	p>0.1	p<0.001	p<0.1
	Forest											
Population	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Selection Criteria	p<0.1	p<0.001	p<0.001	p<0.001	p<0.1	p<0.001	p<0.001	p<0.001	p<0.1	p<0.001	p<0.001	p<0.001
Heritability	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
SI Weighting	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001	p>0.1	p>0.1	p<0.001	p<0.001
Heritability : Selection Criteria	p>0.1	p>0.1	p<0.1	p<0.001	p>0.1	p>0.1	p>0.1	p<0.001	p>0.1	p>0.1	p>0.1	p<0.001
Heritability : SI Weighting	p>0.1	p>0.1	p<0.001	p>0.1	p>0.1	p>0.1	p<0.001	p<0.1	p>0.1	p>0.1	p<0.001	p<0.1

BLUEs of unit per unit cost in z_1 gain and variance targets exhibit large differences by population type, with greater changes occurring in wild population development when compared to landrace and orphan development driven by the near complete fixation of the trait in these populations prior to development phase (Supplemental Figure 11-12). Differences between selection method, narrow-sense heritabilities, and selection index weighting are minimal for this trait within each population type (Supplemental Figure 11-12) and the significant effect of these varies depending on trait target, cropping system, and cost level (Figure 13). However, identifiable differences are found when shifting to BLUEs of unit per unit cost in z_2 gain and variance targets, especially when phenotyping costs increase into horticultural and forestry (Supplemental Figure 13-14). For example, once phenotyping costs are much greater than single sample genotyping costs, such as under forestry crops, the benefits of GS become apparent towards increasing unit gain per unit cost compared to the PRS baseline through significant effect of selection criteria (Figure 13). Under field crop situations, across the range of costs PRS maintains a slight edge over GS in performance of unit gain where GS begins to form a strategic edge over PRS in horticultural crops. Once phenotyping costs increase under horticultural and forestry crops, the unit loss of variance per unit cost is massive under GS with PRS and MxAv maintaining more variation.

Figure 14: Boxplot of baseline scaled trait target unit by unit cost change during population development. The simple trait targets are panels (A) and (B), representing gain and variance change per unit cost, respectively. The complex trait targets are panels (C) and (D), representing gain and variance change per unit cost, respectively.



The cropping system and the population type play a major role in the selection of a scheme for population development. The benefits of GS towards unit gain per unit cost increases moving from field to horticultural, and again to forestry (Supplemental Figure 13). As expected, these benefits towards gain have equally large effects on variation loss, increasingly detrimental moving through these same cropping systems. However, if the target is fixation of z_1, which landrace and orphan populations typically possess prior to population development, then GS is most cost-effective method towards fixation in the wild population with PRS functioning similarly but requiring high narrow-sense heritability (Supplemental Figure 12, Figure 14). Shifting to maintaining variation in z_2, MxAv minimizes unit variance loss per unit cost across all population types and cropping systems, with larger differences found under the more costly agricultural systems (Supplemental Figure 14). Moreover, decreasing narrow-sense heritability and index weighting of z_2 could prevent variance reduction and champion sustainable improvement of the trait (Figure 14).

4.3.2 Best linear unbiased predictions of breeding population sizes

Equally important to other parameters is the size of the breeding population and the effective size of the population from which germplasm is acquired. These random effects of number of parents, number of progeny per cross, and starting effective population size show varying effects through increasing numbers as well as across and within breeding target by cost combinations. For example, on average the parameters that generate the desired effect from the population average vary by the target over the baseline method: (1) z_1 gain unit per unit cost are 20 parents per cycle (16.03), 5 progeny per cross (14.81), and 100 effective population size of germplasm acquisition (10.71); (2) z_1 reduction in variance unit per unit cost are 10 parents per cycle (-9.12), 1 progeny per cross (-8.08), and 100 effective population size (-2.84); (3) z_2 gain unit per unit cost are 25 or 10 parents per cycle (0.04), 20 progeny per cross (0.03), and 25 effective population size (0.02); and (4) z_2 increase in variance unit per unit cost are 20 parents per cycle (0.03), 10 progeny per cross (0.03), and 50 effective population size (0.03). Moreover, within each random effect trends are observed. Effective population size of 100 is an important component during neo-domestication to meet every goal for the specific targets (z_1 gain increase, z_1 variance reduction, z_2 gain increase, and z_2 variance increase). The number of parents and progeny per cross have differential effects depending on the specific target, including 5 progeny per cross meeting the goals of z_1 gain and variance while staying neutral towards z_2 goals and 25 parents per cycle meeting the goals of gain in z_1 and z_2 but adversely affecting the variance goals for each trait. Moreover, the benefit or detriment of each parameters' random effect will improve or worsen when shifting crop-types from field to horticultural to forestry crops, with the largest benefit or detriment being found in the low cost of each crop-type.

4.3.3 Cycle asymptotes of target change

Landrace and orphan population types have a lower asymptote on the logistic or sigmoidal curve for variance prior to population development phase ($cycle < 40$). When fixation of z_1 is the target, wild populations benefit from increased narrow-sense heritability and high weighting of selection index, with the asymptote found for PRS and GS at cycle 45, and MxAv only one cycle later. However, the upper asymptote on the sigmoidal curve for phenotypic

gain of z_2 is greatest across population types under MxAv selection when the trait has low heritability (Supplemental Figure 15). Under MxAv selection, wild and orphan populations approach the asymptote of phenotypic gain (z_2), on average, around cycle 75 where landrace populations approach the asymptote 10 cycles sooner.

4.4 Discussion

Fixation of the simple, cultivation constraining trait is of utmost importance within neo-domestication breeding programs. Successful modern domestication has been demonstrated through genetic engineering (Zhu and Zhu, 2021); however, these methods require years of foundational research and are only accessible to developed breeding programs and domestication candidates related to major crop species (Van Tassel et al., 2020). Our findings show that phenotypic recurrent selection (PRS) is very fast, if heritability for the trait is large, with the lower asymptote of variation found only 5 cycles of selection into population development. However, deleterious variation is always of concern during the rapid reduction of population sizes in breeding, an occurrence observed during historical domestication using comparative genomics (Morrell et al., 2012), where maximum avoidance (MxAv) or alternative mate allocation methods may leave more variation for more complex traits while still approaching the asymptote of simple trait variation rapidly (6 cycles of selection). New mutations are always forming; however, the standing load of deleterious variation exceeds the rate at which the new mutations arise. Once these variants are present in the crop breeding population as standing variation, the genetic bottlenecks of domestication and breeding allow them to drift to higher frequency (Kono et al., 2016, Moyers et al., 2018). Bottlenecks will purge some deleterious mutations (reducing load), but it will also convert masked load into realized load. Prolonged bottlenecks tend to fix deleterious mutations with balance restored using migration (genetic rescue), championing the support for further investigation towards dual population breeding during de novo domestication. As populations shift from large (accumulated masked load) to small populations, load is lost due to random genetic drift and purged by selection with deleterious mutations with large fitness effect becoming exposed due to inbreeding. Demographic bottlenecks can affect the partition of genetic load in different ways, the extent of which is dependent on the length of bottleneck and the effective population size. Increasing the duration of a bottleneck will increase the proportion of realized load to masked load, with total load being less than found during short bottleneck durations (Bertorelle et al., 2022). Therefore, our findings of high starting effective population sizes for breeding (100) meeting target goals is critical towards reducing total load that will be present during incipient domestication given the rapid bottlenecking of the large starting population to the small manageable breeding populations we simulated. In crop improvement, intense selection over short time periods is coupled with reduction in effective population size and limited recombination and followed by migration events and population expansion (MacQueen et al., 2022), returning some balance back to the population following the fixation of beneficial traits, such as z_1, as well as deleterious variants that are also swept through.

The second goal during neo-domestication, often conducted concurrently with the first

and continuing into crop improvement, is maximizing the phenotypic gain of a more complex trait, typically a harvestable organ such as seed, tuber, fruit, or leaf size (Jungers et al., 2023). We show that genomic selection (GS) offers the most cost-effective method towards increasing complex trait gain in half the time of PRS, ranging from 6% more than PRS baseline in landrace population to 53% in wild population. However, sustainability of such an increase is of concern given GS's accuracy reducing trait variation (Jannink, 2010, Olsen and Wendel, 2013). Forfeiting some gain in these traits may be a practical solution when working with more developed germplasm, such as orphan/semi-domestics given their proximity to marketability, where MxAv in our study loses 50% and 20% less variation in z_2 in landrace and orphan populations, respectively. For example, in the scheme of low heritability (0.30) and high index weighting (0.75) for the more complex trait, the difference in gain per unit cost of MxAv compared to GS exhibits a negative relationship with cost, meaning under low cost for field crops GS gain per unit cost is only 0.14 more than MxAv and under high cost for field crops GS is actually 0.03 less than MxAv over the 40 cycles of selection. This specific trend continues through the different crop-types, with the differences in response increasing in horticultural and again in forestry. Moreover, the size of the breeding population has an influence on the complex trait's gain and variation, where more parents and more progeny per cross account for a positive effect per unit of cost. This effect of more progeny per cross is likely derived through an increasing selection intensity, increasing the response to selection (Cobb et al., 2019, Covarrubias-Pazaran et al., 2022). Furthermore, the mean z_2 BLUEs of variance target maintenance finding MxAv with 0.7 heritability and 0.25 weighting to lose ∼20% more variance in wild populations compared to the baseline should be considered because of the beginning of directional selection in the wild population, where the baseline orphan population began this process during burn-in operations (Table 10). The cycle which the asymptote of variance for z_2 in wild is found, on average, at later cycles of selection than compared to MxAv selection in other populations as well as other selection methods across all populations (Supplemental Figure 15), outlining that although the change may be greater, the reduced approach towards complete variance reduction is achieved through MxAv.

Equally important to the target goals of a neo-domestication breeding program is the starting point (i.e. germplasm) from which the program begins. Wild species, including crop wild relatives, possess useful variation for adaptation towards abiotic and biotic stress, novel nutritional complexes, pharmacological uses, and niche human needs (Brozynska et al., 2016, Bailey-Serres et al., 2019). We show through larger populations, both traits' gain per unit cost is increased, alongside a more dramatic bottleneck and loss of variation. This trend is observed primarily through the random effects of starting effective population size and breeding population (crossing) parameters. Over bottlenecking the population may be a temptation, especially when small proportions of individuals contain alleles for simple domestication syndrome traits, but it should be considered that our 100 Ne starting point schemes are already bottlenecked from a wider wild population. There are implications for further increasing this bottleneck during these breeding programs where a continuous necessity to perform genetic rescue/migration may occur (Kono et al., 2016, Bertorelle et al., 2022). However, the target goals shift if considering the crop-type along with its specific agroecosystem, harvestable organ, and primary reproduction. It is also a constantly moving target, where crop uses shift over time and degrading variation in a complex trait early on forces the sustainability of breeding populations towards an inability to adapt to new emerging markets and/or ge-

ographies. Outlined by DeHaan et al. there are specific considerations when starting the domestication of a crop beginning by defining an agricultural target to be met with a type of crop that does not yet exist. Once domesticated, the crop can then be bred for adaptation to different target environments, but only where the variation for the trait exists. Sourcing germplasm from the wild will also require care towards the identification of species, populations, or subpopulations with preadaptation to the agroecosystem which in turn will inform the strategy for domestication of that species (DeHaan et al., 2016). Preadaptations in a wild population may place that breeding population more closely towards our orphan/semi-domestic population schemes, given the proximity to fixation of z_1, and then consideration of maintaining z_2 sustainable gain could be prioritized.

This study attempts the most comprehensive investigation to date in understanding and varying the number of parameters and applying a range of costs towards optimizing breeding schemes during neo-domestication. All possible combinations and parameter iterations were not tested, providing a rich source for future research into stochastic simulation across a wider range of genetic architectures, more traits, new selection methods, and many more. We specifically parameterized given the incipient domestication work currently conducted at the Hawaii Agriculture Research Center across a wide range of crop-types, each with different target goals. The broad applicability of this work can be seen when compared against neo-domestication efforts in intermediate wheatgrass, silphium, pennycress, the gene-editing work in ground cherry, and ancient domesticates (DeHaan et al., 2018, Vilela et al., 2018, Sedbrook et al., 2014, Mueller et al., 2017, 2019). The reported 23% increase in seed size in intermediate wheatgrass is similar to our 30% simulated increase in z_2 under PRS (DeHaan et al., 2018), further matching the reported 30% increase in biomass of silphium (Vilela et al., 2018). We know these gains are possible and practical, but the goal of neo-domestication is not gain for gain's sake, it is to enrich regional food systems through sustainability and variety. Proper market analyses should be conducted prior to embarking on neo-domestication to ensure a market for the product exists or could eventually exist. It is widely accepted that emerging crops must meet market needs (Runck et al., 2014, DeHaan et al., 2016, Messina et al., 2023). Once a market need is identified and a species with a harvestable organ to fill the need identified, then a neo-domestication program compared to a public sector breeding program is comparatively less considering the achievable gains (Coe et al., 2020). The concern, however, remains in the funding opportunity in the public-sector and meeting market needs for variety release in the private sector.

Synthesis: Conclusions and Recommendations

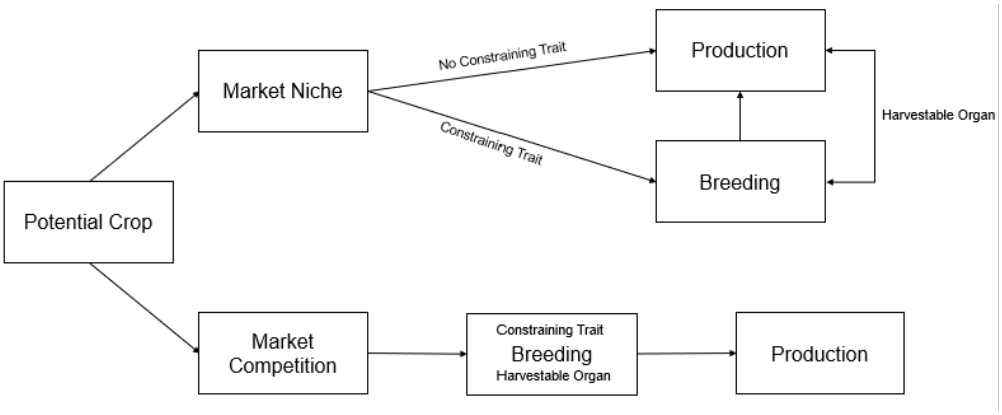
The empirical evidence gathered within this dissertation confirm theory and highlight specific action during neo-domestication and wide hybridization breeding programs. *Acacia koa* was used to better understand crossing and the affect of breeding population size and disease resistance on the potential gain of the domestication syndrome trait early seedling vigor. Elucidated in this chapter was the importance of maintaining large breeding population sizes over the number of progeny, especially during disease susceptibility. *Theobroma cacao* was used to improve our understanding of evaluation through the augmentation of precision of estimates under varied population development and the affect of this change on the potential gain of the domestication syndrome trait seed size. The affect of subsampling is highlighted in this chapter, where max subsampling is pertinent to precise genotypic estimation and appropriate selection in developed breeding populations while reducing subsampling can be leveraged in developing breeding populations to limit costs. *Stevia rebaudiana* was used to differentiate the affect of alternative selection criteria on the important domestication syndrome trait of photoperiod sensitivity. Phenotypic recurrent selection was shown to be effective during early stages of breeding the semi-domesticate, genomic selection increases the potential gain by reducing the length of a cycle time but requires more environments of evaluation, and mate allocation methods can maintain gain for more cycles of selection through reducing the inbreeding coefficient.

The simulated evidence gathered in the final chapter of this dissertation is the most robust analysis to date, looking into $> 5,000$ parameter combinations throughout the breeding cycle and integrating a large range of costs across different cropping systems and population types. The breadth of the study identified specific strategic decisions that can improve breeding program progress towards specific targets during neo-domestication. Germplasm acquisition is a critical moment in any breeding program, but especially so during neo-domestication as larger breeding population size and effective population size beneficially effect the approach to these targets in all schemes, population types, and cropping systems (Supplemental Figure 16). When considering specific targets, the parameter combinations which beneficially effect the approach to that goal change. The fixation of the cultivation contraining simple trait, z_1 in the simulations, is most rapidly approached through mid-level number of breeding parents, low number of progeny per cross, high narrow-sense heritability, and phenotypic recurrent selection using high index weighting (Supplemental Figure 17). Rapid and cost-effective gain in a complex trait, z_2 in the simulations, is maximized using large number of parents and progeny per cross, high narrow-sense heritability, genomic selection and high index weighting (Supplemental Figure 18). However, this will decrease additive genetic variation. Therefore, cost-effective maintenance and sustainable gain in a complex trait is achieved using large number of parents, mid-level number of progeny per cross, low narrow-sense heritability, mate allocation methods with low index weighting (Supplemental Figure 19).

Although this dissertation outlines some strategic methods to begin and conduct neo-domestication breeding programs, the process is long and can be resource intensive. Neo-

domestication for the sake of research is valuable for informing the empirical estimates and fine-tuning strategic options through stochastic simulation, especially as it relates to artificial selection and evolution. However, it is important to consider potential revenue sources from any incipient program by first identifying markets which the newly domesticated crop could fit (Figure 15). If there is a specific niche within the market that a new crop will fit, then the breeder must focus on whether or not there is a constraining trait to cultivation or breeding. When there is not a constraining trait, the crop can likely shift directly into production, exemplified by stevia, where breeding for the improvement of the harvestable organ and cultivar release can be conducted concurrently to production. However, when a constraining trait exists, the program must focus on improving and fixing within the population first. Identifying a market niche for a potential crop with a harvestable organ is of utmost importance and should be considered at the outset of any program. If there is not niche market and the neo-domestication candidate will enter into market competition, then breeding will focus on an index of the cultivation constraining trait and harvestable organ. The candidate crop will not enter into the market until novelty or competitive advantage is shown, either through novel nutritional complexes, ecosystem services, or some other benefit to farmers, consumers, and society.

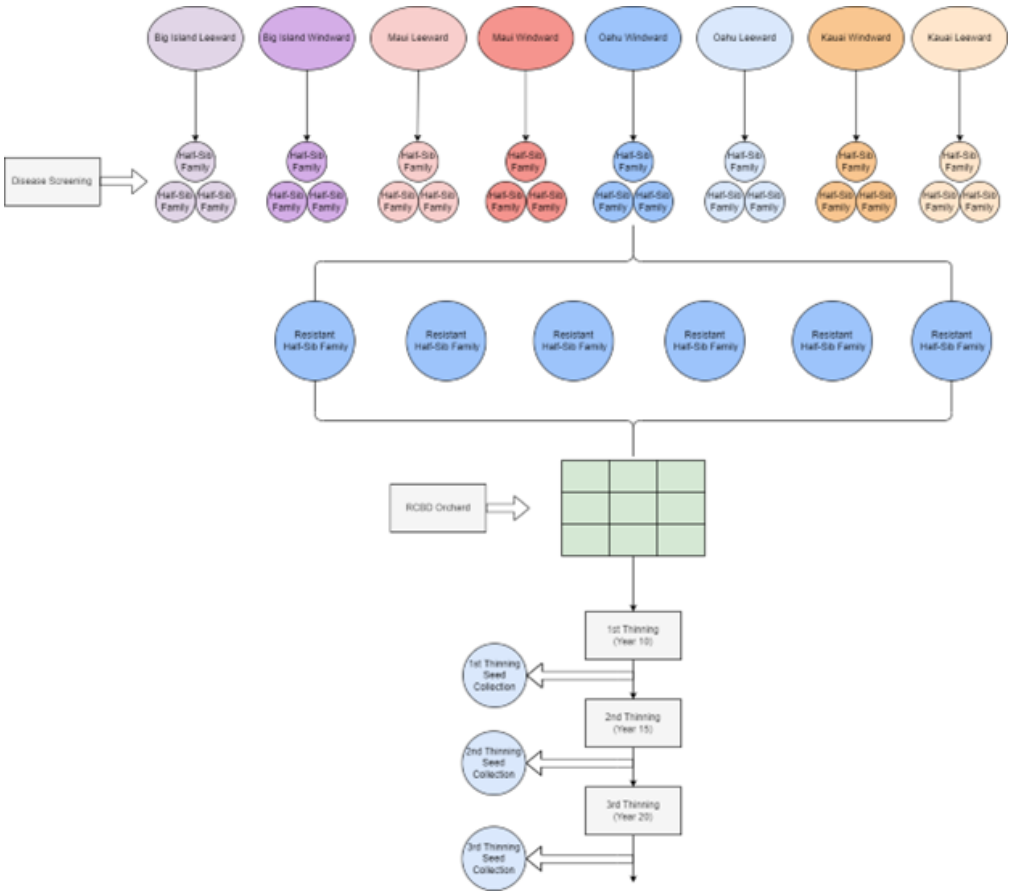
Figure 15: Flow chart outlining the overarching considerations when beginning a neo-domestication breeding program.



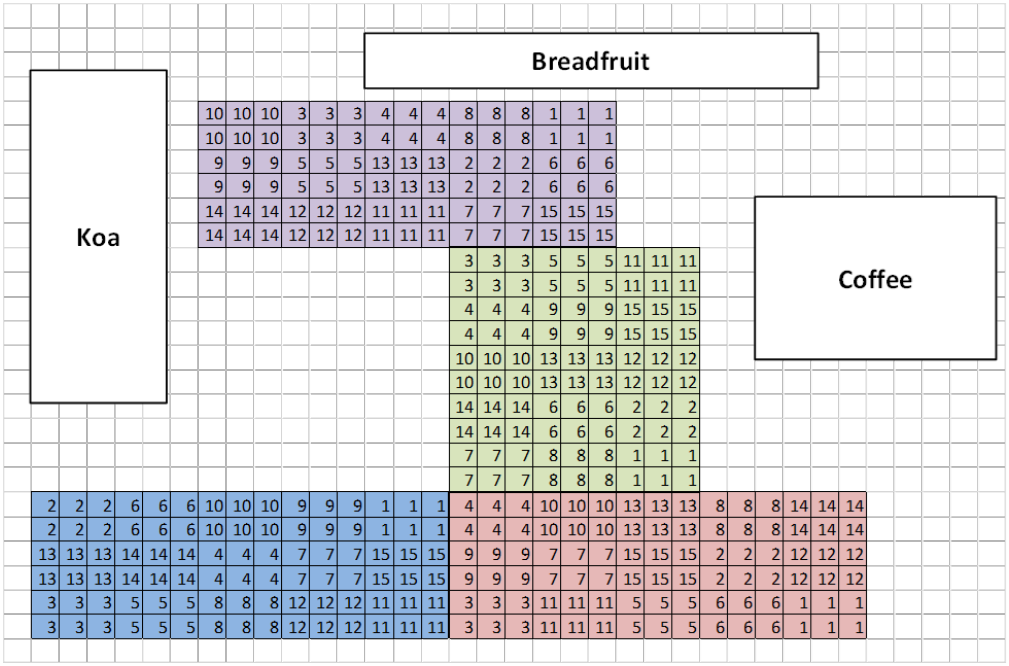
Appendix

Supplemental Figures and Tables

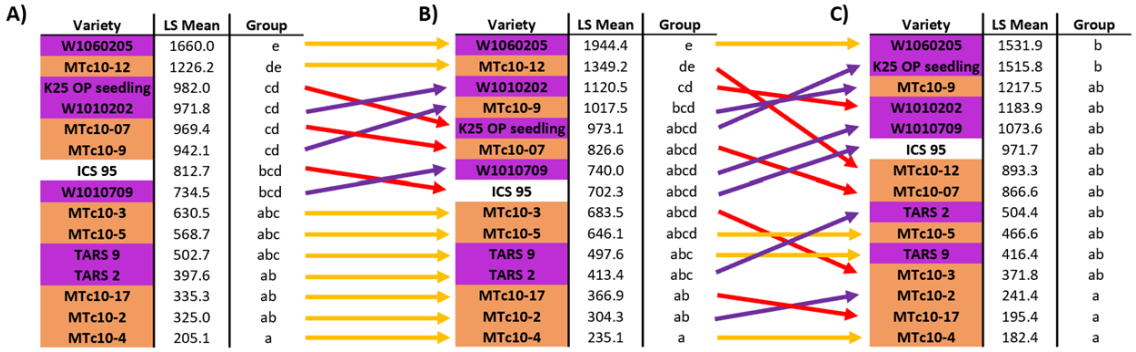
Supplementary Figure 1: Koa wild collection, pedigree, and operational diagram to illustrate the process of deriving thinning group seedling collections.



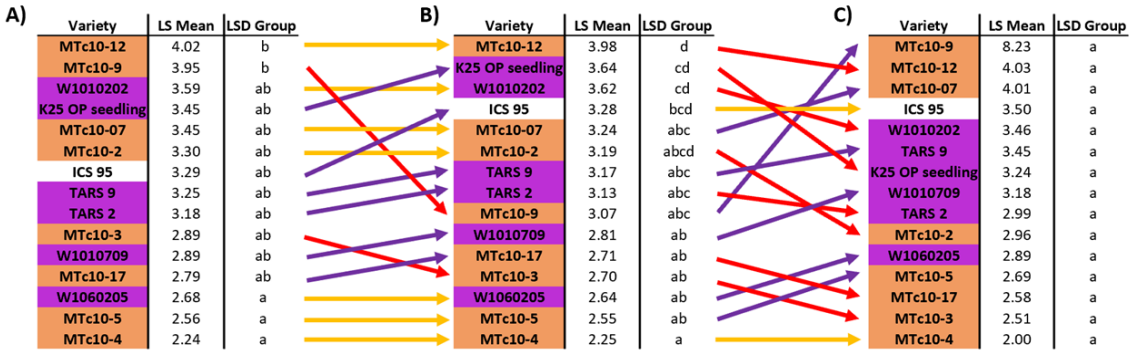
Supplementary Figure 2: Trial map including blocks.



Supplementary Figure 3: Least-Significant Difference rank change across varied sub-sampling for total seed weight (grams) with A) full sub-sampling, B) half sub-sampling, and C) single sample. Arrows represent no rank change (yellow), negative rank change (red), and positive rank change (purple). Varieties a marked as open-pollinated landrace (magenta), hybrid variety (orange), and production control (none).



Supplementary Figure 4: Least-Significant Difference rank change across varied sub-sampling for mean seed size (grams) with A) full sub-sampling, B) half sub-sampling, and C) single sample. Arrows represent no rank change (yellow), negative rank change (red), and positive rank change (purple). Varieties a marked as open-pollinated landrace (magenta), hybrid variety (orange), and production control (none).

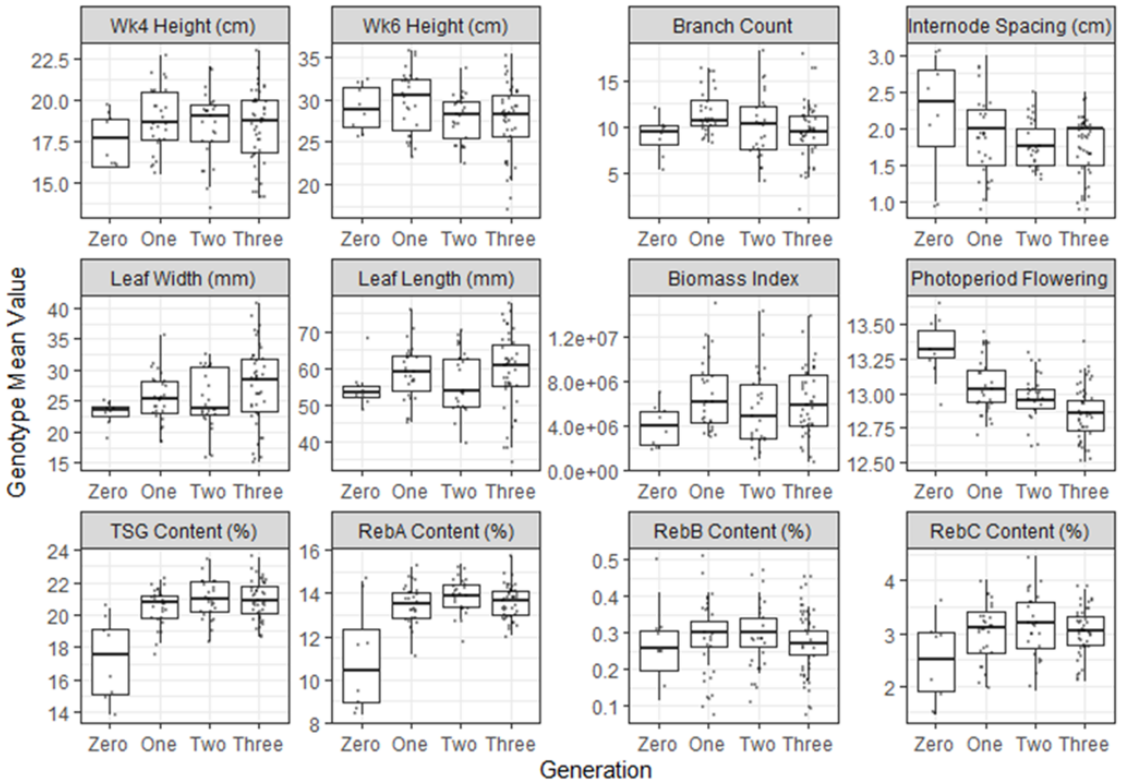


A) Variety	LS Mean	LSD Group
MTc10-12	4.02	b
MTc10-9	3.95	b
W1010202	3.59	ab
K25 OP seedling	3.45	ab
MTc10-07	3.45	ab
MTc10-2	3.30	ab
ICS 95	3.29	ab
TARS 9	3.25	ab
TARS 2	3.18	ab
MTc10-3	2.89	ab
W1010709	2.89	ab
MTc10-17	2.79	ab
W1060205	2.68	a
MTc10-5	2.56	a
MTc10-4	2.24	a

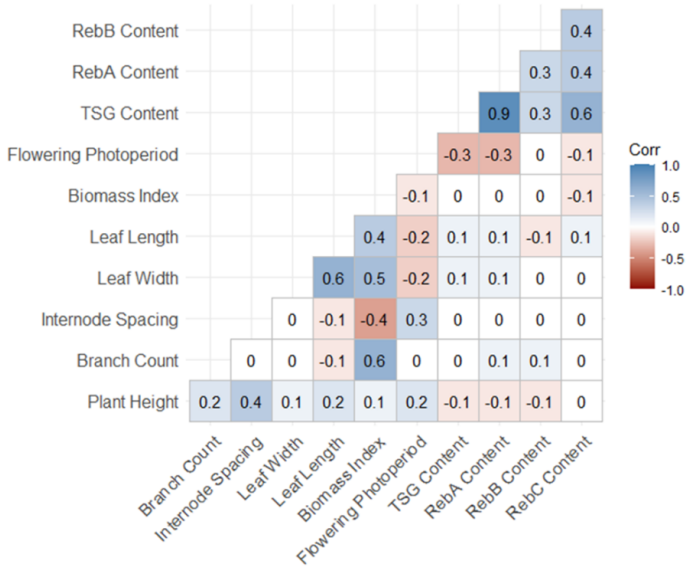
B) Variety	LS Mean	LSD Group
MTc10-12	3.98	d
K25 OP seedling	3.64	cd
W1010202	3.62	cd
ICS 95	3.28	bcd
MTc10-07	3.24	abc
MTc10-2	3.19	abcd
TARS 9	3.17	abc
TARS 2	3.13	abc
MTc10-9	3.07	abc
W1010709	2.81	ab
MTc10-17	2.71	ab
MTc10-3	2.70	ab
W1060205	2.64	ab
MTc10-5	2.55	ab
MTc10-4	2.25	a

C) Variety	LS Mean	LSD Group
MTc10-9	8.23	a
MTc10-12	4.03	a
MTc10-07	4.01	a
ICS 95	3.50	a
W1010202	3.46	a
TARS 9	3.45	a
K25 OP seedling	3.24	a
W1010709	3.18	a
TARS 2	2.99	a
MTc10-2	2.96	a
W1060205	2.89	a
MTc10-5	2.69	a
MTc10-17	2.58	a
MTc10-3	2.51	a
MTc10-4	2.00	a

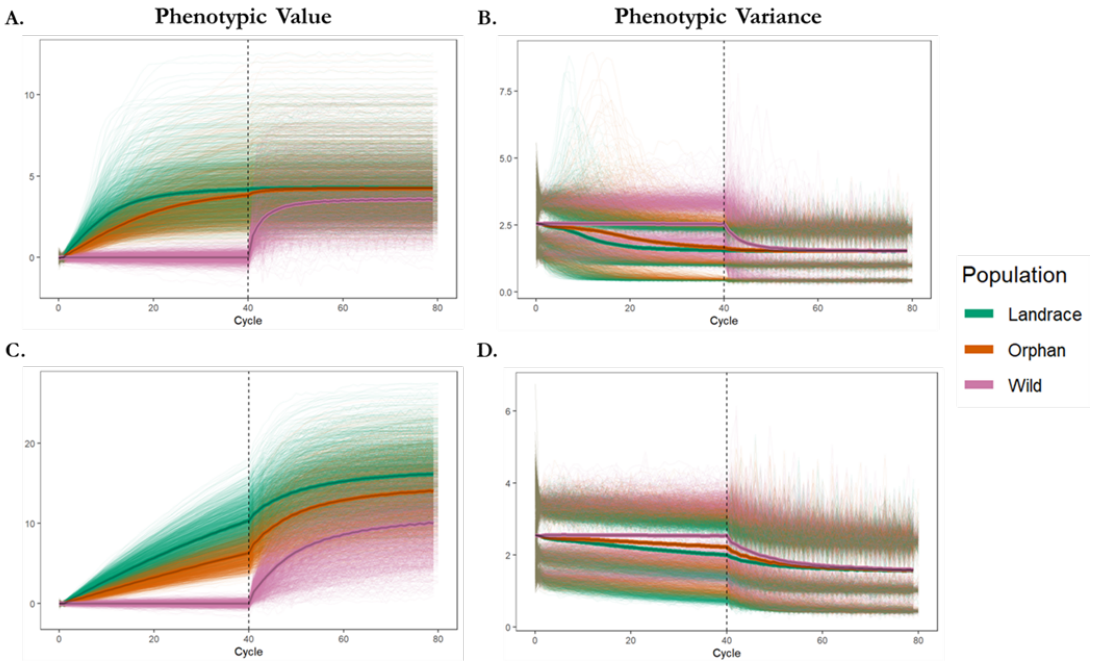
Supplementary Figure 5: Mean trait values observed per genotype during controlled environment trial.



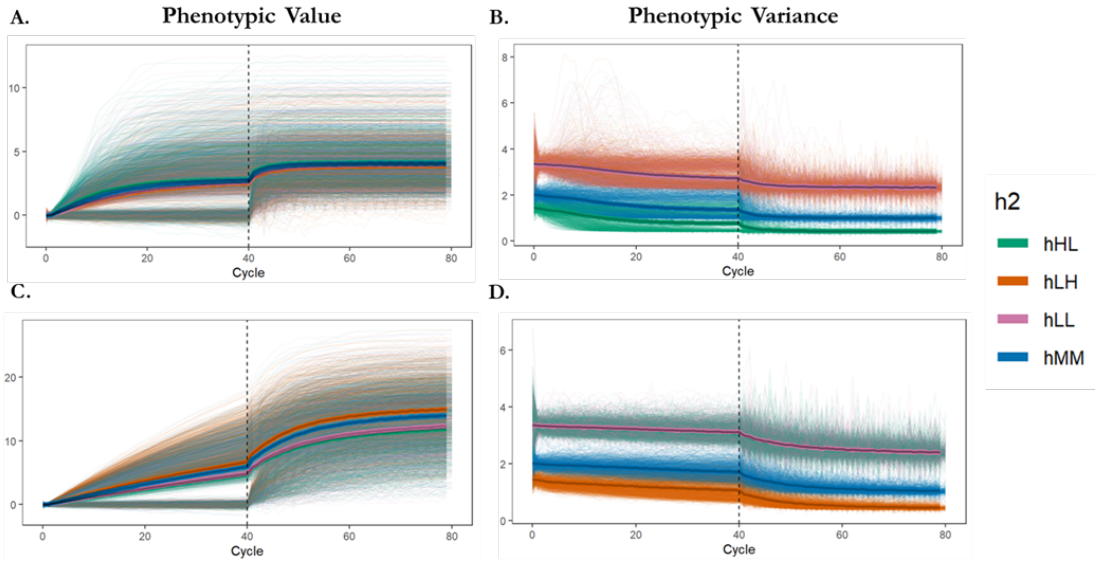
Supplementary Figure 6: Stevia trait correlations.



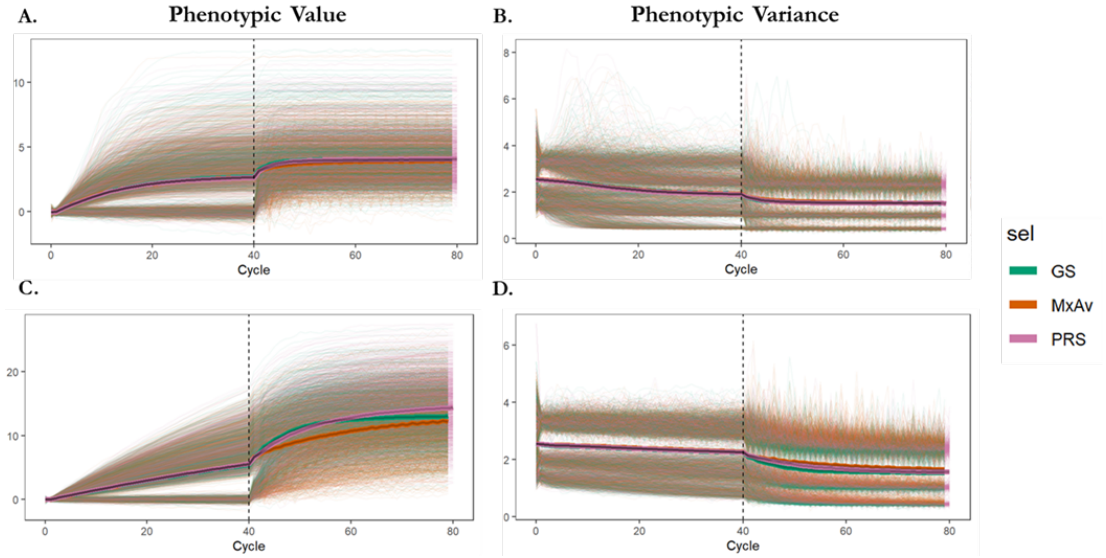
Supplementary Figure 7: All stochastic simulations of phenotypic gain of (A) z_1 and (C) z_2 as well as phenotypic variance of (B) z_1 and (D) z_2 with mean of population type overlayed and grouped by population type.



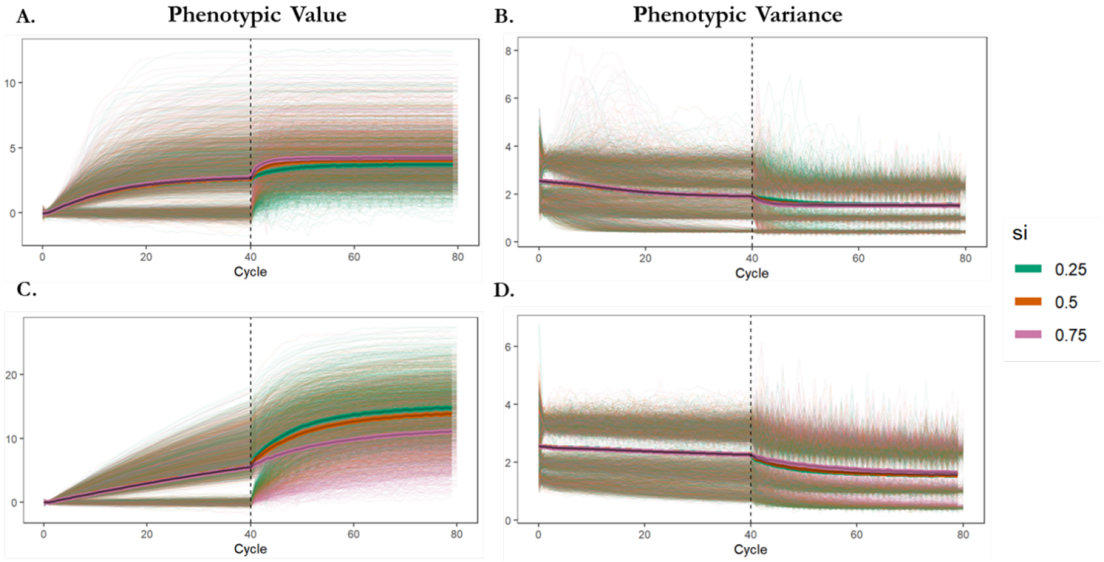
Supplementary Figure 8: All stochastic simulations of phenotypic gain of (A) z_1 and (C) z_2 as well as phenotypic variance of (B) z_1 and (D) z_2 grouped by heritability with mean of narrow-sense heritability overlayed.



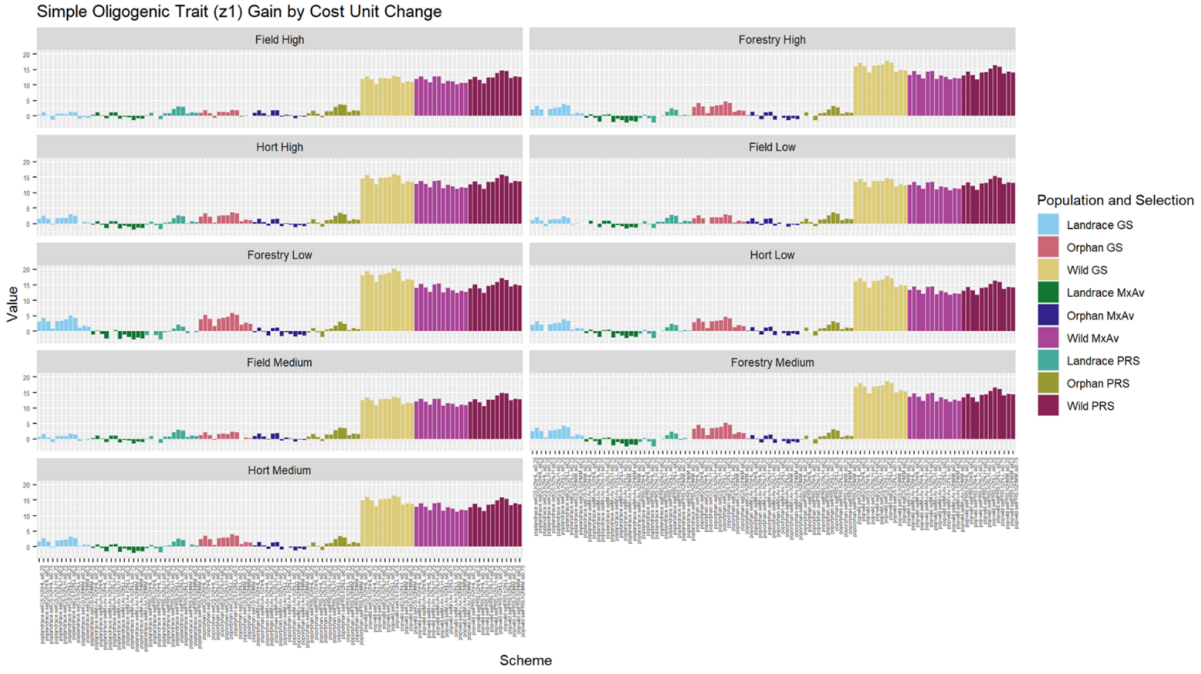
Supplementary Figure 9: All stochastic simulations of phenotypic gain of (A) z_1 and (C) z_2 as well as phenotypic variance of (B) z_1 and (D) z_2 grouped by selection method with mean of each selection method overlayed.



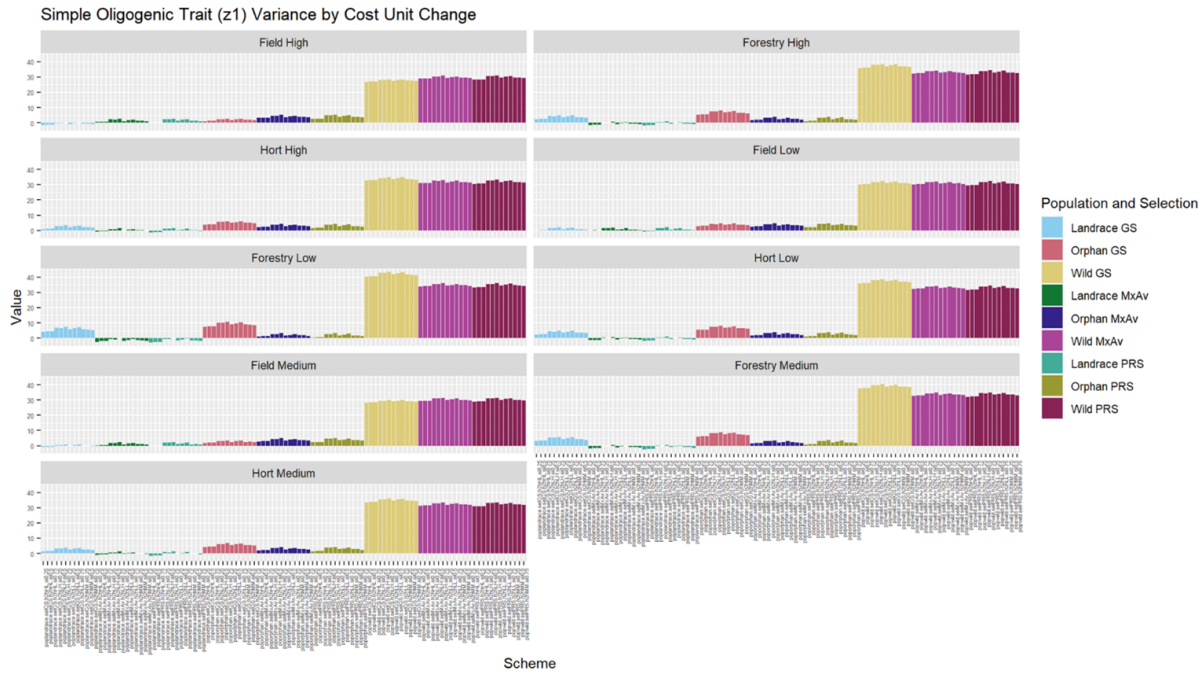
Supplementary Figure 10: All stochastic simulations of phenotypic gain of (A) z_1 and (C) z_2 as well as phenotypic variance of (B) z_1 and (D) z_2 grouped by selection index weighting with mean of each weighting overlayed. The weighting listed is the amount of weight placed on z_1, with the amount of weight placed on z_2 being 1-si weight.



Supplementary Figure 11: Phenotypic gain unit change by unit cost of z_1. X-axis represents the scheme combination of population type, selection method, narrow-sense heritability, and selection index weighting with Y-axis representing the BLUE of unit by unit change, scaled to the baseline scheme.

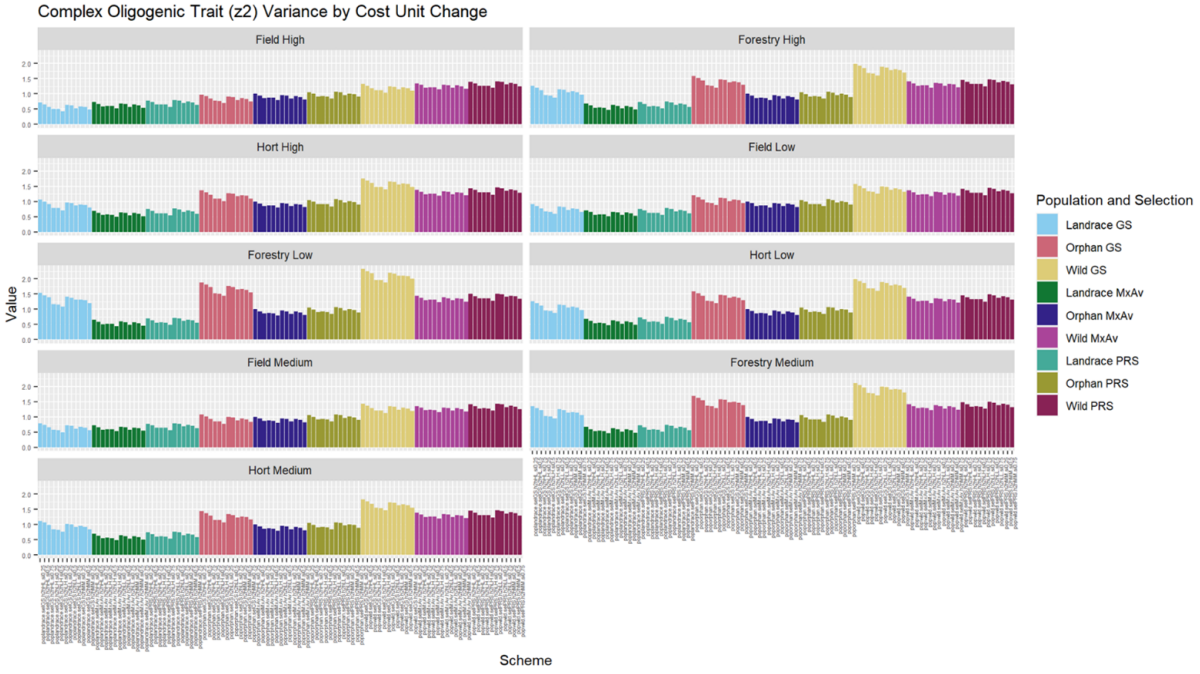


Supplementary Figure 12: Phenotypic variance unit change by unit cost of z_1. X-axis represents the scheme combination of population type, selection method, narrow-sense heritability, and selection index weighting with Y-axis representing the BLUE of unit by unit change, scaled to the baseline scheme.

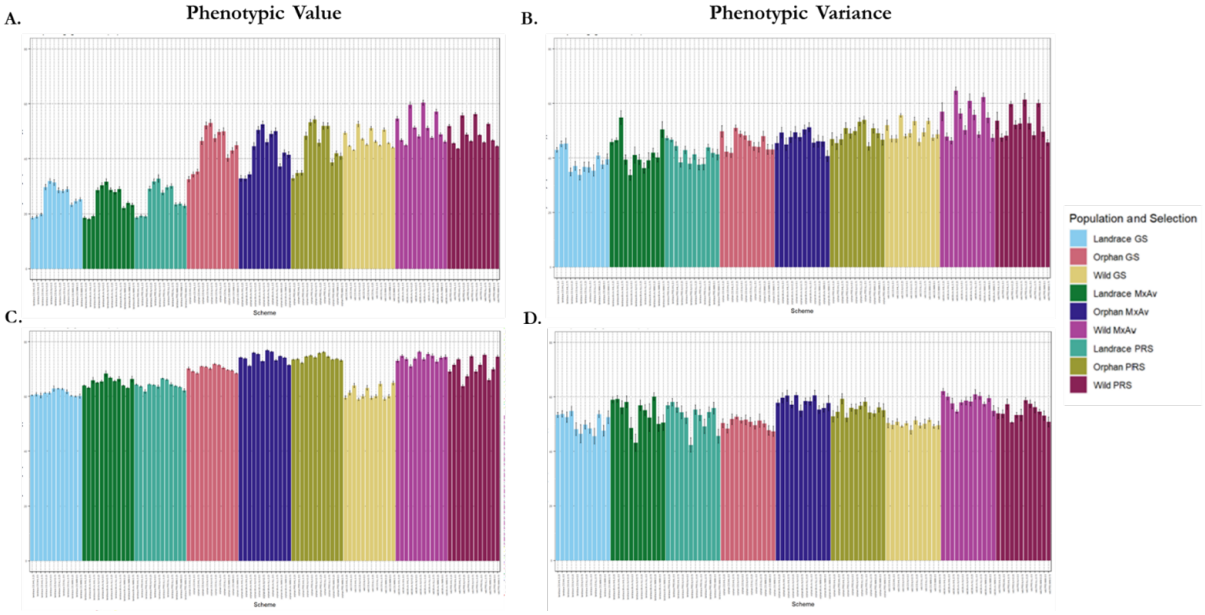


Supplementary Figure 13: Phenotypic gain unit change by unit cost of z_2. X-axis represents the scheme combination of population type, selection method, narrow-sense heritability, and selection index weighting with Y-axis representing the BLUE of unit by unit change, scaled to the baseline scheme.

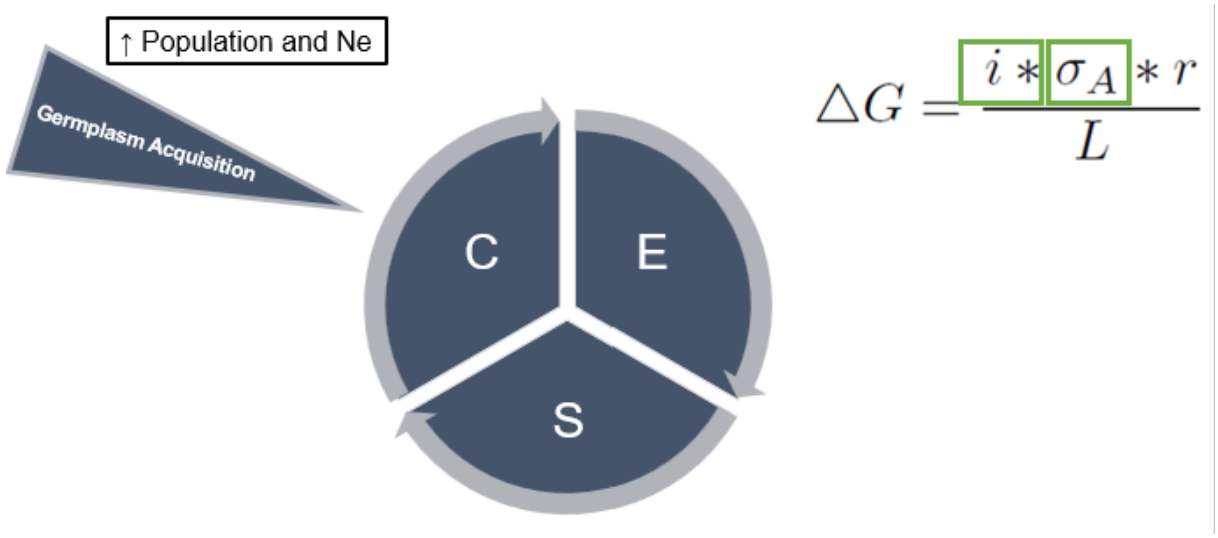
Supplementary Figure 14: Phenotypic variance unit change by unit cost of z_2. X-axis represents the scheme combination of population type, selection method, narrow-sense heritability, and selection index weighting with Y-axis representing the BLUE of unit by unit change, scaled to the baseline scheme.



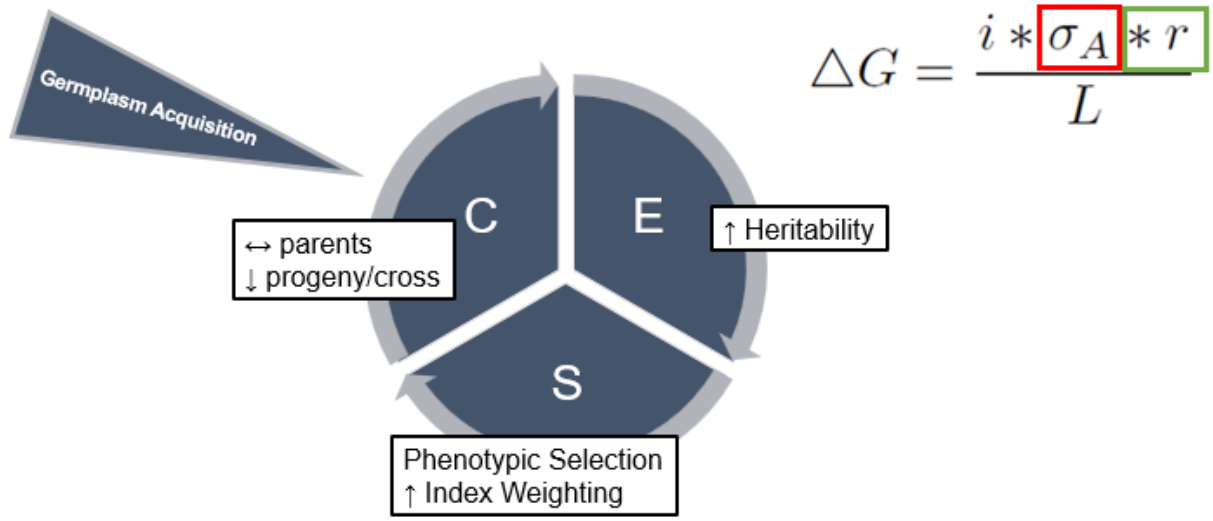
Supplementary Figure 15: Cycle asymptote of target through sigmoidal curve modelling.



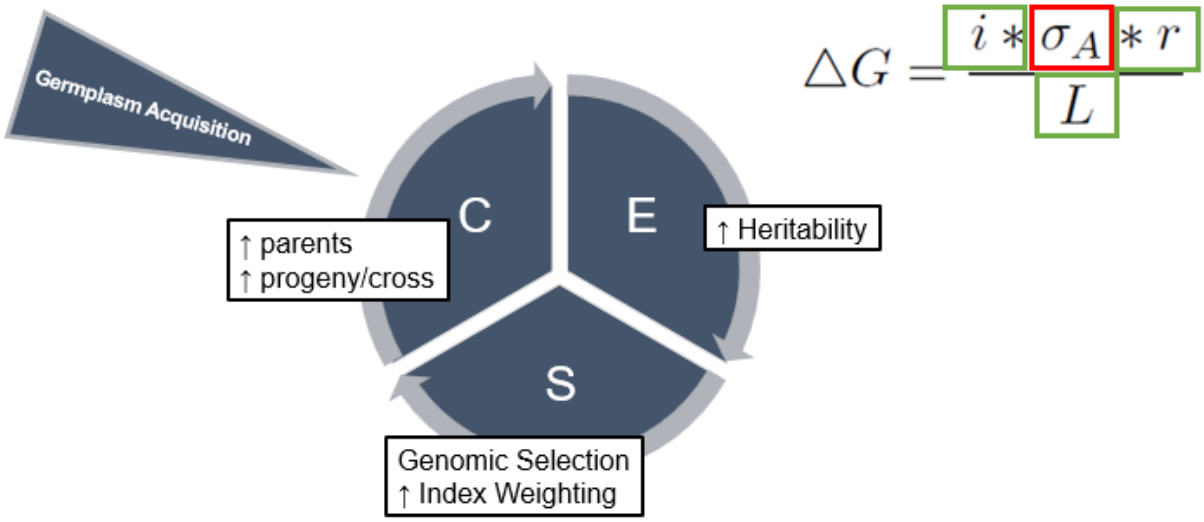
Supplementary Figure 16: Schematic showing the influence of germplasm acquisition at improving the response to selection.



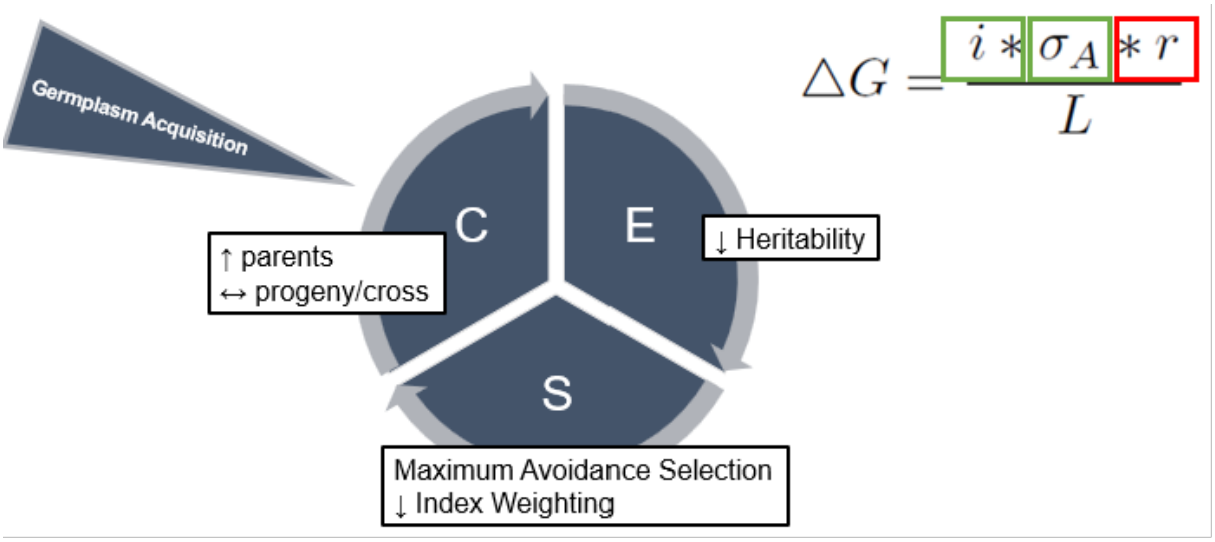
Supplementary Figure 17: Schematic showing the influence of parameter combinations at achieving the target goal of z_1 fixation.



Supplementary Figure 18: Schematic showing the influence of parameter combinations at cost-effective maximization of the response to selection of z_2.



Supplementary Figure 19: Schematic showing the influence of parameter combinations at achieving the target goal of z_2 gain maintenance.



Additional Work

Publications

1. Fumia, N., Nair, R., Lin Y.P., Bishop-von Wettberg, E., Kantar, M.B., Schafleitner, R. (2023). Leveraging genomics and phenomics to speed improvement in mung bean (In Review - The Plant Phenome)

2. Jungers, J, Ewing, PM, Runck, B, Maaz, T, Carlson, C, Neyhart, J, **Fumia, N**, Bajgain, P, Subedi, S, Sharma, S, Senay, S, Hunter, M, Cureton, C, Gutknecht, J, Kantar, M. (2023). Adapting perennial grain and oilseed crops for climate resiliency. Crop Science, https://doi.org/10.1002/csc2.20972.

3. Fumia, N, Kantar, MB, Lin, YP, Schafleitner, R, Lefebvre, V, Paran, I, Börner, A, Diez, MJ, Prohens, J, Bovy, A, Boyaci, F, Pasev, G, Tripodi, P, Barchi, L, Giuliano, G, Barchenger, DW. (2023). Exploration of high-throughput data for heat tolerance selection in Capsicum annuum. The Plant Phenome, https://doi.org/10.1002/ppj2.20071.

4. Fumia, N., Pironon, S., Rubinoff, D., Khoury, C. K., Gore, M. A., Kantar, M. B. (2022). Wild relatives of potato may bolster its adaptation to new niches under future climate scenarios. Food and Energy Security, https://doi.org/10.1002/fes3.360.

5. Fumia, N., Rubinoff, D., Zenil-Ferguson, R., Khoury, C.K., Pironon, S., Gore, M.A., Kantar, M.B. (2022). Interactions between breeding system and ploidy affect niche breadth in Solanum. R. Soc. Open Sci. 9: 211862. https://doi.org/10.1098/rsos.211862.

Presentations

1. Fumia, N. "The International Research Center experience and predictive line selection in mung bean at WorldVeg." An invited seminar presentation at the Graduate Seminar Series in the Tropical Plant and Soil Science Department: UH Manoa TPSS, Honolulu, HI. 8 September 2023.

2. Fumia, N. and Burden, J. "The history of breeding tropical crops at the Hawaii Agriculture Research Center." An invited talk for the UH Manoa program Hoākamai! Building Expertise in FACT Using Active Learning (BE-FACTUAL): Maunawili, HI. 21 July 2023.

3. Fumia, N., Nair, R., Lin Y.P., Bishop-von Wettberg, E., Kantar, M.B., Schafleitner, R. "Leveraging genomics and phenomics to speed improvement in mung bean." A poster presentation at the National Association of Plant Breeders Conference (NAPB): Greenville, SC. 16-20 July 2023.

4. Fumia, N., "Predictive line selection with high-throughput data." An invited workshop training at the WorldVeg Headquarters: Tainan, Taiwan. 30 March 2023.

5. Fumia, N., Lin, Y.P., Shafleitner, R. "Development of a pipeline for line selection with insights into the mung bean mini-core collection." An invited seminar presentation at the

WorldVeg Headquarters: Tainan, Taiwan. 30 March 2023.

6. Fumia, N., Kantar, MB, Lin, YP, Schafleitner, R, Lefebvre, V, Paran, I, Börner, A, Diez, MJ, Prohens, J, Bovy, A, Boyaci, F, Pasev, G, Tripodi, P, Barchi, L, Giuliano, G, Barchenger, DW. "Exploration of High-Throughput Data for Heat Tolerance Selection in Capsicum annuum." A poster presentation at Plant and Animal Genome Conference (PAG30): San Diego, CA. 13-18 January 2023.

7. Fumia, N., Kantar, M.B., Khoury, C. "Wild relatives to potato may bolster its adaptation to new niches under future climate scenarios." An invited webinar presentation to the International Center for Tropical Agriculture (CIAT) – Alliance Biodiversity. 7 June 2022.

8. Fumia, N., Wolfe, M.D., Zenil-Ferguson, R., Kantar, M.B. "Simulation and Evidence: Comparison of predicted and realized phenotypic gain during the domestication of Stevia rebaudiana." A poster presentation at the 46th Annual Tester Symposium (Best Graduate Poster Runner-Up), UH Manoa School of Life Sciences: Honolulu, HI. 20-22 April 2022.

Workshop Training

1. Griffith, E. and Sharp, J. "Navigating Tough Conversations in Statistical Collaboration." University of Hawaii at Manoa, Honolulu, HI. 30 August 2022

2. Bernardo, R. "Genomewide Markers in Plant Breeding." University of Minnesota, Minneapolis-Saint Paul, MN. 22-24 June 2022.

3. Lorenz, A. "Data Bootcamp for Genomic Prediction in Plant Breeding." University of Minnesota, Minneapolis-Saint Paul, MN. 20-22 June 2022.

4. Byrne, D., Riera-Lizarazu, O., Endelman, J. "Tools for Genomics-Assisted Breeding in Polyploids." San Diego State University, San Diego, CA. 13-14 January 2022.

References

D. J. Adamski, N. S. Dudley, C. W. Morden, and D. Borthakur. Genetic differentiation and diversity of acacia koa populations in the hawaiian islands. *Plant Species Biology*, 27(3): 181–190, 2012.

R. G. Allaby, D. Q. Fuller, and T. A. Brown. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences*, 105(37):13982–13986, 2008.

R. G. Allaby, R. L. Ware, and L. Kistler. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evolutionary Applications*, 12(1):29–37, 2019.

A. Allier, S. Teyssèdre, C. Lehermeier, L. Moreau, and A. Charcosset. Optimized breeding strategies to harness genetic resources with different performance levels. *BMC genomics*, 21:1–16, 2020.

R. C. Anderson, D. E. Gardner, C. C. Daehler, and F. C. Meinzer. Dieback of acacia koa in hawaii: ecological and pathological characteristics of affected stands. *Forest Ecology and Management*, 162(2-3):273–286, 2002.

L. G. Angelini, A. Martini, B. Passera, and S. Tavarini. Cultivation of stevia rebaudiana bertoni and associated challenges. *Sweeteners*, pages 35–85, 2018.

J. Bailey-Serres, J. E. Parker, E. A. Ainsworth, G. E. Oldroyd, and J. I. Schroeder. Genetic strategies for improving crop yields. *Nature*, 575(7781):109–118, 2019.

P. J. Baker. *Koa (Acacia koa) ecology and silviculture*, volume 211. United States Department of Agriculture, Forest Service, Pacific Southwest ..., 2009.

J. Bančič, C. R. Werner, R. C. Gaynor, G. Gorjanc, D. A. Odeny, H. F. Ojulong, I. K. Dawson, S. P. Hoad, and J. M. Hickey. Modeling illustrates that genomic selection provides new opportunities for intercrop breeding. *Frontiers in Plant Science*, 12:605172, 2021.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

H. Becker and J. Leon. Stability analysis in plant breeding. *Plant breeding*, 101(1):1–23, 1988.

F. Bekele and W. Phillips-Mora. Cacao (theobroma cacao l.) breeding. In *Advances in plant breeding strategies: Industrial and food crops*, pages 409–487. Springer, 2019.

S. Ben-Sadoun, R. Rincent, J. Auzanneau, F.-X. Oury, B. Rolland, E. Heumez, C. Ravel, G. Charmet, and S. Bouchet. Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theoretical and Applied Genetics*, 133:2197–2212, 2020.

R. Bernardo. *Breeding for quantitative traits in plants*, volume 1. Stemma press Woodbury, MN, 2002.

R. Bernardo. Genomewide selection when major genes are known. *Crop Science*, 54(1):68–75, 2014.

R. Bernardo. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something blue. *Heredity*, 125(6):375–385, 2020.

R. Bernardo and J. Yu. Prospects for genomewide selection for quantitative traits in maize. *Crop Science*, 47(3):1082–1090, 2007.

G. Bertorelle, F. Raffini, M. Bosse, C. Bortoluzzi, A. Iannucci, E. Trucchi, H. E. Morales, and C. van Oosterhout. Genetic load: genomic estimates and applications in non-model animals. *Nature Reviews Genetics*, 23(8):492–503, 2022.

D. A. Bisognin, K. H. Lencina, H. P. Greff, T. Tonetto, and D. Gazzana. Progeny evaluation and early selection for plant height in acacia mearnsii improve genetic gains. *Crop Breeding and Applied Biotechnology*, 22, 2023.

N. Bondarev, M. Sukhanova, O. Reshetnyak, and A. Nosov. Steviol glycoside content in different organs of stevia rebaudiana and its dynamics during ontogeny. *Biologia plantarum*, 47:261–264, 2003.

F. M. Bourland. Functional characterization of seed and seedling vigor in cotton. *The Journal of Cotton Science*, 23:168–176, 2019.

J. Brandle and N. Rosa. Heritability for yield, leaf: stem ratio and stevioside content estimated from a landrace cultivar of stevia rebaudiana. *Canadian journal of plant Science*, 72(4):1263–1266, 1992.

J. Brandle, A. Starratt, and M. Gijzen. Stevia rebaudiana: Its agricultural, biological, and chemical properties. *Canadian Journal of plant science*, 78(4):527–536, 1998.

G. K. Brown, D. J. Murphy, J. Kidman, and P. Y. Ladiges. Phylogenetic connections of phyllodinous species of acacia outside australia are explained by geological history and human-mediated dispersal. *Australian Systematic Botany*, 25(6):390–403, 2012.

M. Brozynska, A. Furtado, and R. J. Henry. Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant biotechnology journal*, 14(4):1070–1085, 2016.

M. B. Burke, D. B. Lobell, and L. Guarino. Shifts in african crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Global Environmental Change*, 19(3):317–325, 2009.

G. D. Carr. Chromosome numbers of hawaiian flowering plants and the significance of cytology in selected taxa. *American journal of botany*, 65(2):236–242, 1978.

D. Chapman, B. V. Purse, H. E. Roy, and J. M. Bullock. Global trade networks determine the distribution of invasive non-native species. *Global Ecology and Biogeography*, 26(8): 907–917, 2017.

E. Cheesman. Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture*, 21(8), 1944.

G. K. Chen, P. Marjoram, and J. D. Wall. Fast and flexible simulation of dna sequence data. *Genome research*, 19(1):136–142, 2009.

C. R. Clement, D. Cristo-Araújo, G. Coppens D'Eeckenbrugge, A. Alves Pereira, D. Picanço-Rodrigues, et al. Origin and domestication of native amazonian crops. *Diversity*, 2(1): 72–106, 2010.

C. Clemente, L. G. Angelini, R. Ascrizzi, and S. Tavarini. Stevia rebaudiana (bertoni) as a multifunctional and sustainable crop for the mediterranean climate. *Agriculture*, 11(2): 123, 2021.

J. N. Cobb, G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, 126(4):867–887, 2013.

J. N. Cobb, R. U. Juma, P. S. Biswas, J. D. Arbelaez, J. Rutkoski, G. Atlin, T. Hagen, M. Quinn, and E. H. Ng. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theoretical and applied genetics*, 132:627–645, 2019.

M. T. Coe, K. M. Evans, K. Gasic, and D. Main. Plant breeding capacity in us public institutions. *Crop Science*, 60(5):2373–2385, 2020.

P. Cosson, C. Hastoy, L. E. Errazzu, C. J. Budeguer, P. Boutié, D. Rolin, and V. Schurdi-Levraud. Genetic diversity and population structure of the sweet leaf herb, stevia rebaudiana b., cultivated and landraces germplasm assessed by est-ssrs genotyping and steviol glycosides phenotyping. *BMC plant biology*, 19(1):1–11, 2019.

G. Covarrubias-Pazaran. Genome-assisted prediction of quantitative traits using the r package sommer. *PloS one*, 11(6):e0156744, 2016.

G. Covarrubias-Pazaran, Z. Gebeyehu, D. Gemenet, C. Werner, M. Labroo, S. Sirak, P. Coaldrake, I. Rabbi, S. I. Kayondo, E. Parkes, et al. Breeding schemes: what are they, how to formalize them, and how to improve them? *Frontiers in Plant Science*, 12:791859, 2022.

L. DeHaan, M. Christians, J. Crain, and J. Poland. Development and evolution of an intermediate wheatgrass domestication program. *Sustainability*, 10(5):1499, 2018.

L. R. DeHaan, D. L. Van Tassel, J. A. Anderson, S. R. Asselin, R. Barnes, G. J. Baute, D. J. Cattani, S. W. Culman, K. M. Dorn, B. S. Hulke, et al. A pipeline strategy for grain crop domestication. *Crop Science*, 56(3):917–930, 2016.

H. Dempewolf, G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino. Past and future use of wild relatives in crop breeding. *Crop science*, 57(3):1070–1082, 2017.

J. Dudley and R. Moll. Interpretation and use of estimates of heritability and genetic variances in plant breeding 1. *Crop science*, 9(3):257–262, 1969.

N. Dudley, R. L. James, and A. Yeh. *Comparative virulence of Hawaiian Fusarium oxysporum isolates on Acacia koa seedlings*. US Department of Agriculture, Forest Service, Northern Region, Forest Health . . . , 2007.

N. Dudley, T. Jones, R. James, R. Sniezko, J. Wright, C. Liang, P. F. Gugger, and P. Cannon. Applied genetic conservation of hawaiian acacia koa: An eco-regional approach. In *In: Sniezko, Richard A.; Man, Gary; Hipkins, Valerie; Woeste, Keith; Gwaze, David; Kliejunas, John T.; McTeague, Brianna A., tech. cords. 2017. Gene conservation of tree species—banking on the future. Proceedings of a workshop. Gen. Tech. Rep. PNW-GTR-963. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station: 78-91.*, volume 963, pages 78–91, 2017.

N. Dudley, T. Jones, K. Gerber, A. L. Ross-Davis, R. A. Sniezko, P. Cannon, and J. Dobbs. Establishment of a genetically diverse, disease-resistant acacia koa a. gray seed orchard in kokee, kauai: Early growth, form, and survival. *Forests*, 11(12):1276, 2020.

N. S. Dudley, T. C. Jones, R. L. James, R. A. Sniezko, P. Cannon, and D. Borthakur. Applied disease screening and selection program for resistance to vascular wilt in hawaiian acacia koa. *Southern Forests: A Journal of Forest Science*, 77(1):65–73, 2015.

A. Eskes and C. Lanaud. Cocoa. 2001.

A. Eyre-Walker, R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences*, 95(8):4441–4446, 1998.

D. Falconer and T. Mackay. Introduction to quantitative genetics. essex. *UK: Longman Group*, 12, 1996.

A. R. Fernie and J. Yan. De novo domestication: an alternative route toward new crops for the future. *Molecular plant*, 12(5):615–631, 2019.

K. Finlay and G. Wilkinson. The analysis of adaptation in a plant-breeding programme. *Australian journal of agricultural research*, 14(6):742–754, 1963.

S. Flint-Garcia, M. J. Feldmann, H. Dempewolf, P. L. Morrell, and J. Ross-Ibarra. Diamonds in the not-so-rough: Wild relative diversity hidden in crop genomes. *PLoS biology*, 21(7): e3002235, 2023.

J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, et al. Solutions for a cultivated planet. *Nature*, 478(7369):337–342, 2011.

C. Fonseca, D. Viands, J. Hansen, and A. Pell. Associations among forage quality traits, vigor, and disease resistance in alfalfa. *Crop science*, 39(5):1271–1276, 1999.

D. Q. Fuller. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the old world. *Annals of Botany*, 100(5):903–924, 2007.

D. Q. Fuller, T. Denham, M. Arroyo-Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby, and M. D. Purugganan. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences*, 111(17):6147–6152, 2014.

W. C. Gagne. Canopy-associated arthropods in acacia koa and metrosideros tree communities along an altitudinal transect on hawaii island. *Pacific insects*, 21(1):56–82, 1979.

D. E. Gardner et al. Acacia koa seedling wilt caused by fusarium oxysporum f. sp. koae, f. sp. nov. *Phytopathology*, 70(7):594–597, 1980.

B. S. Gaut, C. M. Díez, and P. L. Morrell. Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in genetics*, 31(12):709–719, 2015.

R. C. Gaynor, G. Gorjanc, A. R. Bentley, E. S. Ober, P. Howell, R. Jackson, I. J. Mackay, and J. M. Hickey. A two-part strategy for using genomic selection to develop inbred lines. *Crop Science*, 57(5):2372–2386, 2017.

R. C. Gaynor, G. Gorjanc, and J. M. Hickey. Alphasimr: an r package for breeding program simulations. *G3*, 11(2):jkaa017, 2021.

G. Gorjanc, R. C. Gaynor, and J. M. Hickey. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theoretical and applied genetics*, 131:1953–1966, 2018.

S. C. Grossnickle and J. E. MacDonald. Why seedlings grow: influence of plant attributes. *New forests*, 49:1–34, 2018.

J. R. Harlan, J. De Wet, and E. G. Price. Comparative evolution of cereals. *Evolution*, 27 (2):311–325, 1973.

J. R. Harlan et al. *Crops and man.* American Society of Agronomy, 1975.

L. Hazel and J. L. Lush. The efficiency of three methods of selection. *Journal of Heredity*, 33(11):393–399, 1942.

L. N. Hazel. The genetic basis for constructing selection indexes. *Genetics*, 28(6):476–490, 1943.

C. R. Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pages 423–447, 1975.

J. M. Hickey, S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B. M. Prasanna, M. Grondona, A. Zambelli, V. S. Windhausen, K. Mathews, et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, 54(4):1476–1488, 2014.

J. M. Hickey, T. Chiurugwi, I. Mackay, and W. Powell. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature genetics*, 49(9):1297–1303, 2017.

S. Hotchkiss and J. O. Juvik. A late-quaternary pollen record from ka ‘au crater, o ‘ahu, hawai ‘i. *Quaternary Research*, 52(1):115–128, 1999.

A. Iwaro, F. Bekele, and D. Butler. Evaluation and utilisation of cacao (theobroma cacao l.) germplasm at the international cocoa genebank, trinidad. *Euphytica*, 130(2):207–221, 2003.

J.-L. Jannink. Dynamics of long-term genomic selection. *Genetics Selection Evolution*, 42 (1):1–11, 2010.

J.-L. Jannink, R. Astudillo, and P. Frazier. Insight into a two-part plant breeding scheme through bayesian optimization of budget allocations. *agriRxiv*, (2023):20230277076, 2023.

S. Jansky, H. Dempewolf, E. L. Camadro, R. Simon, E. Zimnoch-Guzowska, D. Bisognin, and M. Bonierbale. A case for crop wild relative preservation and use in potato. *Crop Science*, 53(3):746–754, 2013.

J. D. Johnson and M. L. Cline. Seedling quality of southern pines. In *Forest regeneration manual*, pages 143–159. Springer, 1991.

J. Jungers, B. Runck, P. M. Ewing, T. Maaz, C. Carlson, J. Neyhart, N. Fumia, P. Bajgain, S. Subedi, V. Sharma, et al. Adapting perennial grain and oilseed crops for climate resiliency. *Crop Science*, 2023.

L. Kaplan. Archeology and domestication in american phaseolus (beans). *Economic Botany*, 19:358–368, 1965.

N. Khan, L. Motilal, D. Sukha, F. Bekele, A. Iwaro, G. Bidaisee, P. Umaharan, L. Grierson, and D. Zhang. Variability of butterfat content in cacao (theobroma cacao l.): combination and correlation with other seed-derived traits at the international cocoa genebank, trinidad. *Plant Genetic Resources*, 6(3):175–186, 2008.

K. Khoshbakht and K. Hammer. How many plant species are cultivated? *Genetic resources and crop evolution*, 55:925–928, 2008.

M. Kimura and J. F. Crow. On the maximum avoidance of inbreeding. *Genetics Research*, 4(3):399–415, 1963.

A. D. Kinghorn, D. Soejarto, N. Nanayakkara, C. Compadre, H. Makapugay, J. Hovanec-Brown, P. Medon, and S. Kamath. A phytochemical screening procedure for sweet ent-kaurene glycosides in the genus stevia. *Journal of natural products*, 47(3):439–444, 1984.

J. Kofsky, H. Zhang, and B.-H. Song. Genetic architecture of early vigor traits in wild soybean. *International Journal of Molecular Sciences*, 21(9):3105, 2020.

T. J. Kono, F. Fu, M. Mohammadi, P. J. Hoffman, C. Liu, R. M. Stupar, K. P. Smith, P. Tiffin, J. C. Fay, and P. L. Morrell. The role of deleterious substitutions in crop genomes. *Molecular biology and evolution*, 33(9):2307–2317, 2016.

M. J. Kovach, M. T. Sweeney, and S. R. McCouch. New insights into the history of rice domestication. *TRENDS in Genetics*, 23(11):578–587, 2007.

P. Lachenaud and G. Oliver. Variability and selection for morphological bean traits in wild cocoa trees (theobroma cacao l.) from french guiana. *Genetic Resources and Crop Evolution*, 52(3):225–231, 2005.

B. Laliberté, N. Cryer, A. J. Daymond, M. End, J. M. Engels, A. Eskes, M. Gilmour, P. Lachenaud, W. Phillips-Mora, C. J. Turnbull, et al. A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy. 2012.

D. Laloë. Precision and information in linear models of genetic evaluation. *Genetics Selection Evolution*, 25(6):557–576, 1993.

R. Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, pages 314–334, 1976.

R. Lande. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, pages 402–416, 1979.

R. Lass and G. A. R. Wood. *Cocoa production: present constraints and priorities for research*. The World Bank, 1985.

J. J. Le Roux, D. Strasberg, M. Rouget, C. W. Morden, M. Koordom, and D. M. Richardson. Relatedness defies biogeography: the tale of two island endemics (acacia heterophylla and a. koa). *New Phytologist*, 204(1):230–242, 2014.

J. Lee, K. Kang, and H. Park. New high rebaudioside-a stevia variety" suweon 11". *The Research Reports of the Office of Rural Development (Korea R.)*, 1982.

C. Lesk, P. Rowhani, and N. Ramankutty. Influence of extreme weather disasters on global crop production. *Nature*, 529(7584):84–87, 2016.

W. H. Lewis. Early uses of stevia rebaudiana (asteraceae) leaves as a sweetener in paraguay. *Economic botany (USA)*, 1992.

A. J. Lorenz. 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, 2013.

J. L. Lush. Progeny test and individual performance as indicators of an animal's breeding value. *Journal of Dairy Science*, 18(1):1–19, 1935.

J. L. Lush et al. Heritability of quantitative characters in farm animals. *Heritability of quantitative characters in farm animals.*, 1949.

M. Lynch. The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution*, 45(3):622–629, 1991.

M. Lynch, B. Walsh, et al. Genetics and analysis of quantitative traits. 1998.

I. Mackay, E. Ober, and J. Hickey. Gpluse: beyond genomic selection. *Food and energy security*, 4(1):25–35, 2015.

A. H. MacQueen, C. K. Khoury, P. Miklas, P. E. McClean, J. M. Osorno, B. C. Runck, J. W. White, M. B. Kantar, and P. M. Ewing. Local to continental-scale variation in fitness and heritability in common bean. *Crop Science*, 62(2):767–779, 2022.

M. Matsui, K. Matsui, Y. Kawasaki, Y. Oda, T. Noguchi, Y. Kitagawa, M. Sawada, M. Hayashi, T. Nohmi, K. Yoshihira, et al. Evaluation of the genotoxicity of stevioside and steviol using six in vitro and one in vivo mutagenicity assays. *Mutagenesis*, 11 (6):573–579, 1996.

Z. Mehrabi, S. Pironon, M. Kantar, N. Ramankutty, and L. Rieseberg. Shifts in the abiotic and biotic environment of cultivated sunflower under future climate change. *OCL*, 26:9, 2019.

C. Messina, L. Borras, T. Tang, and M. Cooper. Crop improvement can accelerate agriculture adaptation to societal demands and climate change. *bioRxiv*, pages 2023–09, 2023.

R. S. Meyer and M. D. Purugganan. Evolution of crop species: genetics of domestication and diversification. *Nature reviews genetics*, 14(12):840–852, 2013.

R. S. Meyer, A. E. DuVal, and H. R. Jensen. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, 196(1):29–48, 2012.

D. J. Midmore and A. H. Rank. *A new rural industry-Stevia-to replace imported chemical sweeteners*. Rural Industries Research and Development Corporation, 2002.

P. L. Morrell, E. S. Buckler, and J. Ross-Ibarra. Crop genomics: advances and applications. *Nature Reviews Genetics*, 13(2):85–96, 2012.

J. C. Motamayor, A.-M. Risterucci, P. A. Lopez, C. F. Ortiz, A. Moreno, and C. Lanaud. Cacao domestication i: the origin of the cacao cultivated by the mayas. *Heredity*, 89(5): 380–386, 2002.

J. C. Motamayor, A.-M. Risterucci, M. Heath, and C. Lanaud. Cacao domestication ii: progenitor germplasm of the trinitario cacao cultivar. *Heredity*, 91(3):322–330, 2003.

J. C. Motamayor, P. Lachenaud, J. W. Da Silva e Mota, R. Loor, D. N. Kuhn, J. S. Brown, and R. J. Schnell. Geographic and genetic population differentiation of the amazonian chocolate tree (theobroma cacao l). *PloS one*, 3(10):e3311, 2008.

B. T. Moyers, P. L. Morrell, and J. K. McKay. Genetic costs of domestication and improvement. *Journal of Heredity*, 109(2):103–116, 2018.

N. G. Mueller, G. J. Fritz, P. Patton, S. Carmody, and E. T. Horton. Growing the lost crops of eastern north america's original agricultural system. *Nature plants*, 3(7):1–5, 2017.

N. G. Mueller, A. White, and P. Szilagyi. Experimental cultivation of eastern north america's lost crops: Insights into agricultural practice and yield potential. *Journal of Ethnobiology*, 39(4):549–566, 2019.

K. M. Olsen and J. F. Wendel. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual review of plant biology*, 64:47–70, 2013.

D. Pauli, P. Andrade-Sanchez, A. E. Carmo-Silva, E. Gazave, A. N. French, J. Heun, D. J. Hunsaker, A. E. Lipka, T. L. Setter, R. J. Strand, et al. Field-based high-throughput plant phenotyping reveals the temporal patterns of quantitative trait loci associated with stress-responsive traits in cotton. *G3: Genes, Genomes, Genetics*, 6(4):865–879, 2016.

M. A. Peixoto, I. F. Coelho, K. A. Leach, L. L. Bhering, and M. F. Resende Jr. Simulation based decision making and implementation of tools in hybrid crop breeding pipelines. *Crop Science*, 2023.

L. Pejchar, K. D. Holl, and J. L. Lockwood. Hawaiian honeycreeper home range size varies with habitat: implications for native acacia koa forestry. *Ecological Applications*, 15(3): 1053–1061, 2005.

S. Pironon, T. R. Etherington, J. S. Borrell, N. Kühn, M. Macias-Fauria, I. Ondo, C. Tovar, P. Wilkin, and K. J. Willis. Potential adaptive strategies for 29 sub-saharan crops under future climate change. *Nature Climate Change*, 9(10):758–763, 2019.

M. D. Purugganan. Evolutionary insights into the nature of plant domestication. *Current Biology*, 29(14):R705–R714, 2019.

M. D. Purugganan and D. Q. Fuller. The nature of selection during plant domestication. *Nature*, 457(7231):843–848, 2009.

R Core Team. R: A language and environment for statistical computing. 2021. URL https://www.R-project.org/.

N. Ramankutty, Z. Mehrabi, K. Waha, L. Jarvis, C. Kremen, M. Herrero, and L. H. Rieseberg. Trends in global agricultural land use: implications for environmental health and food security. *Annual review of plant biology*, 69:789–815, 2018.

J. Ramirez-Villegas and C. K. Khoury. Reconciling approaches to climate change adaptation for colombian agriculture. *Climatic Change*, 119:575–583, 2013.

M. D. V. d. Resende and R. S. Alves. Statistical significance, selection accuracy, and experimental precision in plant breeding. *Crop Breeding and Applied Biotechnology*, 22, 2022.

B. C. Runck, M. B. Kantar, N. R. Jordan, J. A. Anderson, D. L. Wyse, J. O. Eckberg, R. J. Barnes, C. L. Lehman, L. R. DeHaan, R. M. Stupar, et al. The reflective plant breeding paradigm: A robust system of germplasm development to support strategic diversification of agroecosystems. *Crop Science*, 54(5):1939–1948, 2014.

P. Schmidt, J. Hartung, J. Bennewitz, and H.-P. Piepho. Heritability in plant breeding on a genotype-difference basis. *Genetics*, 212(4):991–1008, 2019.

J. C. Sedbrook, W. B. Phippen, and M. D. Marks. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (thlaspi arvense l.). *Plant Science*, 227:122–132, 2014.

X. Shi. Genetic improvement of leucaena spp. and acacia koa gray as high-value hardwoods, 2003.

S. Shizhen. A study on good variety selection in stevia rebaudiana. *Sci. Agric. Sin*, 28:37–41, 1995.

Y. Shyu et al. Effects of harvesting dates on the characteristics, yield, and sweet. *Journal of Agricultural Research of China*, 43(1):29–39, 1994.

L. L. Sloat, S. J. Davis, J. S. Gerber, F. C. Moore, D. K. Ray, P. C. West, and N. D. Mueller. Climate adaptation by crop migration. *Nature communications*, 11(1):1243, 2020.

H. F. Smith. A discriminant function for plant selection. *Annals of eugenics*, 7(3):240–250, 1936.

P. Smỳkal, M. N. Nelson, J. D. Berger, and E. J. Von Wettberg. The impact of genetic changes during crop domestication. *Agronomy*, 8(7):119, 2018.

D. D. Soejarto. Botany of stevia and stevia rebaudiana. In *Stevia*, pages 31–52. CRC Press, 2001.

D. B. South, J. L. Rakestraw, and G. A. Lowerts. Early gains from planting large-diameter seedlings and intensive management are additive for loblolly pine. *New Forests*, 22:97–110, 2001.

W. Sun, J. Brewbaker, and M. Austin. Acacia koa genetic improvement. *Proceedings, Hawai'i Agriculture: Positioning for Growth. CTAHR, University of Hawaii at Manoa*, 1996.

E. Thomas, M. van Zonneveld, J. Loo, T. Hodgkin, G. Galluzzi, and J. van Etten. Present spatial diversity patterns of theobroma cacao l. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS One*, 7(10): e47676, 2012.

I. Valio and R. F. Rocha. Effect of photoperiod and growth regulator on growth and flowering of stevia rebaudiana bertoni. *Japanese Journal of Crop Science*, 46(2):243–248, 1977.

D. L. Van Tassel, O. Tesdell, B. Schlautman, M. J. Rubin, L. R. DeHaan, T. E. Crews, and A. Streit Krug. New food crop domestication in the age of gene editing: genetic, agronomic and cultural change remain co-evolutionarily entangled. *Frontiers in Plant Science*, 11:789, 2020.

A. Vilela, L. González-Paleo, K. Turner, K. Peterson, D. Ravetta, T. E. Crews, and D. Van Tassel. Progress and bottlenecks in the early domestication of the perennial oilseed silphium integrifolium, a sunflower substitute. *Sustainability*, 10(3):638, 2018.

D. Villalba, B. Díez-Unquera, A. Carrascal, A. Bernués, and R. Ruiz. Multi-objective simulation and optimisation of dairy sheep farms: Exploring trade-offs between economic and environmental outcomes. *Agricultural Systems*, 173:107–118, 2019.

E. von Wettberg, T. M. Davis, and P. Smỳkal. Wild plants as source of new crops, 2020.

K. P. Voss-Fels, M. Cooper, and B. J. Hayes. Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics*, 132:669–686, 2019.

W. L. Wagner, D. R. Herbst, and S. H. Sohmer. *Manual of the flowering plants of Hawai'i*. University of Hawaii Press, 1990.

E. Wall, G. Simm, and D. Moran. Developing breeding schemes to assist mitigation of greenhouse gas emissions. *Animal*, 4(3):366–376, 2010.

J. G. Wallace, E. Rodgers-Melnick, and E. S. Buckler. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual review of genetics*, 52:421–444, 2018.

J. Wang, M. Van Ginkel, D. Podlich, G. Ye, R. Trethowan, W. Pfeiffer, I. H. DeLacy, M. Cooper, and S. Rajaram. Comparison of two breeding strategies by computer simulation. *Crop Science*, 43(5):1764–1773, 2003.

C. A. Wartha and A. J. Lorenz. Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities. *Crop Breeding and Applied Biotechnology*, 21, 2021.

A. Watson, S. Ghosh, M. J. Williams, W. S. Cuddy, J. Simmonds, M.-D. Rey, M. Asyraf Md Hatta, A. Hinchliffe, A. Steed, D. Reynolds, et al. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature plants*, 4(1):23–29, 2018.

A. K. Yadav, S. Singh, D. Dhyani, and P. S. Ahuja. A review on the improvement of stevia [stevia rebaudiana (bertoni)]. *Canadian Journal of Plant Science*, 91(1):1–27, 2011.

J. F. Yanagida, J. B. Friday, P. Illukpitiya, R. J. Mamiit, and Q. Edwards. Economic value of hawai 'i's forest industry in 2001. 2004.

L. B. Zaidan, S. M. Dietrich, and G. Felippe. Effect of photoperiod on flowering and stevioside content in plants of stevia rebaudiana bertoni. *Japanese Journal of Crop Science*, 49(4): 569–574, 1980.

J. Zhang, H. Yu, and J. Li. De novo domestication: retrace the history of agriculture to design future crops. *Current Opinion in Biotechnology*, 81:102946, 2023.

X. Zhang, A. Sallam, L. Gao, T. Kantarski, J. Poland, L. R. DeHaan, D. L. Wyse, and J. A. Anderson. Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. *The Plant Genome*, 9(1): plantgenome2015–07, 2016.

Q. Zhu, X. Zheng, J. Luo, B. S. Gaut, and S. Ge. Multilocus analysis of nucleotide variation of oryza sativa and its wild relatives: severe bottleneck during domestication of rice. *Molecular biology and evolution*, 24(3):875–888, 2007.

X.-G. Zhu and J.-K. Zhu. Precision genome editing heralds rapid de novo domestication for new crops. *Cell*, 184(5):1133–1134, 2021.

D. Zohary and P. Spiegel-Roy. Beginnings of fruit growing in the old world: Olive, grape, date, and fig emerge as important bronze age additions to grain agriculture in the near east. *Science*, 187(4174):319–327, 1975.

D. Zohary, M. Hopf, et al. *Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley.* Number Ed. 3. Oxford university press, 2000.