DISSERTATION

GENOMICS OF FLOWERING TIME TO ACCELERATE BREEDING OF DROUGHT-TOLERANT PEARL MILLET FOR SENEGAL

Submitted by

Safiétou Tooli Fall

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Geoffrey Morris

John Mckay
Cristiana Argueso
Esten Mason

ABSTRACT

ELITE INFERENCE IN PEARL MILLET BREEDING GUIDED BY TRAIT ARCHITECTURE AND FLOWERING TIME GENETICS

Distinct selection strategies are required for two categories of traits targeted by breeding programs. While directional selection increases the mean of a desired trait (e.g., grain yield), stabilizing selection is necessary to maintain the optimal state of an acquired trait (e.g., flowering time). Balancing these strategies requires an optimized breeding framework that enables the de novo creation of elite gene pools to guide cross-design and accelerate genetic gain, especially in under-resourced programs. This thesis hypothesizes that trait-informed elite definition, grounded in trait architecture, enhances the precision of breeding programs. The Chapter 1 frames the challenge of managing multiple traits under different selection pressures in traditional and emerging millet programs. The Chapter 2 develops and compares molecular inference strategies to define cis- and trans-elite types. We show that the QGI-based similarity method outperforms other approaches, particularly for known QTL regions. Importantly, this finding highlights that uncovering the genetic basis of key traits is critical to applying this framework effectively. To address this, Chapters 3 and 4 focus on dissecting the genetic control of flowering time, a central adaptive trait in pearl millet. In Chapter 3, a forward genetics approach using genome-wide association studies (GWAS) in West African germplasm revealed an oligogenic architecture, with key loci including *Phytochrome C (PhyC)* driving ecotypic divergence. Population structure

analysis further indicated the existence of shared gene pools at the regional level, offering opportunities for collaborative breeding. In Chapter 4, a reverse genetics approach was used to characterize the gene regulatory network underlying flowering. Expression profiling of early (Souna) and late (Sanio) genotypes showed that flowering is regulated by differential activation of the *GI–CaHd3a* pathway, modulated by photoreceptors such as *CaPhyC* and *CaPhyA*, leading to divergent regulatory dynamics between ecotypes. Together, these results provide the genomic and regulatory basis for trait-informed elite inference and support a breeding strategy that simultaneously conserves elite backgrounds while introducing desired traits. This integrative framework, which leverages population structure, trait architecture, and molecular regulation, demonstrates how quantitative, molecular, and functional genetics can be combined to enhance selection strategies in pearl millet and similar crops facing climate-related challenges.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support, guidance, and inspiration of numerous individuals and institutions, to whom I am deeply grateful.

First and foremost, I would like to express my sincere gratitude to the *Centre d'Étude Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS-ISRA)* for selecting me for a fellowship under the *Sorghum and Millet Innovation Lab (SMIL)* program. This opportunity marked a significant turning point in my academic journey and has been instrumental in the successful completion of my research.

I am profoundly thankful to my co-supervisors, Dr. Geoffrey Morris, Head of the Crop Adaptation Lab at Colorado State University (CSU), and Dr. Ndjido Kane, Director of CERAAS, for their expert guidance, continuous support, and generous mentorship. Their leadership and encouragement shaped not only the direction of my research but also my approach to collaborative science.

I am also deeply grateful to my thesis committee members, Dr. John McKay, Dr. Cristiana Argueso, and Dr. Esten Mason, for their expertise in ecological genomics, molecular biology, and plant breeding, respectively. Their insights greatly enriched the scientific depth and practical relevance of this work.

My experience at CSU was especially enriching thanks to the warm welcome I received at the Crop Adaptation Lab. I extend my sincere thanks to Dr. Jacques Martin Faye and Dr. Fanna Maina, whose mentorship, collaboration, and friendship meant a great deal to me. I am equally

grateful to all the lab members who shared their knowledge and created a dynamic and inspiring work environment.

I gratefully acknowledge the financial and logistical support of the *Feed the Future Sorghum and Millet Innovation Lab (SMIL)* and the *Innovation Lab for Crop Improvement (ILCI)* programs of USAID, which made every phase of this research possible. Special thanks go to Timothy Dalton, Nat Bascom, Kira Everhart-Valentin, Benjamin Kohl, and the entire SMIL and ILCI management teams for their unwavering commitment and support in making my PhD journey a reality.

My heartfelt appreciation goes to the research teams at ISRA/CERAAS and CNRA, particularly Dr. Ghislain Kanfany, who welcomed me with openness and provided an environment conducive to rigorous scientific work. I also thank his entire technical team, especially Mr. Yagouba Diao, who played a key role in coordinating and executing the field experiments linked to this research.

I wish to express my deep gratitude to those who inspired me early in my academic path. Mr. Serigne Mbacke Diop, my biology teacher at Lycée Thierno Seydou Nourou Tall, was the first to awaken my interest in science. Professor Ibrahima Ndoye, Professor of Microbiology and my academic mentor, has consistently supported and encouraged me throughout my journey.

I am honored to acknowledge Professor Mame Ourey Sy, Full Professor of Plant Biology; Professor Mame Thioro Diaw Sow, Associate Professor of Plant Biology; Dr. Yacine Ndour Badiane, former Director of the Laboratory for Research in Nutrition and Plant Production (LRNPV); and Dr. Amadou Alpha Sall, Scientific Director of the Institut Pasteur of Dakar. Their generous decision to introduce me to biotechnology immediately after I completed high school and their continued support and guidance throughout played a pivotal role in shaping both my scientific aspirations and my career path. This dissertation is the result of a collective journey. To everyone who believed in me, supported me, and inspired me this achievement is also yours.

PREFACE

This dissertation reflects the culmination of my doctoral research on the genetic improvement of pearl millet in Senegal. I hold a B.Eng. in Biotechnology from the Université de Sherbrooke (2015) and a Master's in Plant and Microbial Biology from Université Cheikh Anta Diop (2017). My research lies at the intersection of plant breeding, genomics, and crop adaptation to climate stress.

Two chapters of this dissertation have been published:

- *Fall, S.T., et al.* (2023). Genetic Improvement of Pearl Millet in Senegal: Past, Present, and Future Prospects. In *Crop Adaptation and Improvement for Drought-Prone Environments*. Link

- *Fall, S.T., Kena, A., Rice, B.R., Kanfany, G., Diatta, C., Kane, N.A., Fritz, A.K., Morris, G.P.* (2025). Genomic Approaches to Build *de novo* Elite Breeding Gene Pools from Locally-Adapted Landraces. Link

This work contributes to efforts to strengthen locally adapted, genomics-informed breeding strategies for crops in West Africa.

DEDICATION

I dedicate this work to the special people who have made this journey a memorable one.

To my wonderful father, Abdou Salam, who named me after the late distinguished scientist Dr. Safiétou Touré Fall, former Director of Research in Animal Science at ISRA. He offered me not just a name but a role model and a powerful message that women can excel in science and lead by example. His guidance and unwavering support made my scientific journey not only possible but natural. Thank you, Papa.

To my lovely mother, Ndeye Mayé, whose strength, devotion, and loving care allowed me to focus on my work while she cared for my children during my time abroad. Your presence, love, and sacrifice gave me the peace of mind and stability I needed to succeed. Thank you, Maman.

To my dear husband, Ousmane, thank you for your unfailing support, love, and belief in me. You stood by me with patience and encouragement at every stage of this journey.

To my children who endured my absence during important moments in their young lives. Your love and understanding have been my greatest source of strength and motivation.

To my sisters, les Tooli, thank you for being a constant source of support, for your listening ears, your compassion and for stepping in so selflessly to help with my children whenever and wherever needed. Your solidarity has meant the world to me.

Finally, I extend heartfelt thanks to my extended family, my family in-law through Mamadou Ba Sr, my dear brother Mara and my friends, whose encouragement, presence, and kindness have surrounded me every step of the way. Each of you contributed in your own way to the successful completion of this work.

I am deeply grateful for you all

TABLE OF CONTENTS

Chapter 1- Genetics improvement of pearl millet in Senegal: Past, Present and Future Prospects

1.1 Summary

Pearl millet is an important cereal crop for smallholder farmers' food security in Senegal and is grown on approximately 800,000 ha of land with a total national production of 600,000 tons annually. However, its production has been affected by increasing biotic and abiotic stresses such as downy mildew disease, drought and heat, and use of unimproved varieties in farmers' fields. Genetic improvement of pearl millet started in Senegal nearly a century ago and experienced three major phases that were defined by different improvement targets. Over time, Progressive national yield improvements have been made. This chapter summarizes the breeding progress of the Senegal national breeding program in the past century. It also formulates future directions for further improvement of the productivity of pearl millet to meet the increasing national demand for food and nutrition security in the midst of climate change. This chapter posits that an ambitious pearl millet improvement strategy that integrates the modern genetic and genomic tools and novel genetic resources for resistance to biotic and abiotic stresses with variety development work is urgently needed to revolutionize pearl millet production in Senegal.

1.2 Introduction

Pearl millet [*Pennisetum glaucum* (L.) R. Br. syn. *Cenchrus americanus (L.)*] is a diploid (2x = 2n = 14) C4 and allogamous crop that belongs to the family of Poaceae (Bono & Leclercq, 1963). Archaeological and genetic findings have inferred PM was domesticated about 4,900 years ago in the western Sahara belt, more precisely in Northern Mali (Burgarella et al., 2018; Manning et al., 2011; Oumar et al., 2008).

Pearl millet is currently the sixth most important cereal crop in the world following maize (*Zea mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), barley (*Hordium vulgare*), and sorghum (*Sorghum bicolor*) based on harvested area (FAOSTAT, 2019). It accounts for more than half of global millet production and is mainly cultivated in West Africa and India for household food security and income generation, whereas it is cultivated mostly as forage in Brazil, the United States, Mexico, and Australia. West Africa and India cover more than 95% of its global production (FAOSTAT, 2019). India is the largest pearl millet producer, followed by West African countries including Nigeria, Niger, Burkina Faso, Chad, Mali, and Senegal.

In Sub-Saharan Africa, pearl millet was historically grown by farmers for their basic subsistence. Now, it has also become a commercial crop and the main source of income and nutrients in the most drought and heat-stressed production areas. It is one of the rare cereals of traditional farming systems that is able to withstand frequent and prolonged periods of drought and high temperature in saline soils, at both the vegetative and reproductive phases, which would cause crop failure or greater yield reduction in other cereals (Dussert et al., 2015; Serba et al., 2017). Compared to some other domesticated cereals, pearl millet shows a large genetic diversity and an exceptional potential of adaptation to a wide range of ecological conditions, especially drought, across the semiarid areas (Sehgal et al., 2015; Serba et al. 2019).

Another major asset of pearl millet is its potential to relieve malnutrition in developing countries. Food insecurity remains the major issue to be solved in Sub-Saharan Africa, where 22% of the population is severely food insecure (ANSD, 2018). Despite improvements over the last decade, the current food insecurity situation in Senegal remains critical, and 16.5% of the population was undernourished in 2018 (ANSD, 2018). The food demand will increase dramatically in the coming years following continued population growth, with the projected West African population (920 million in mid-2014) doubling by 2050. The Senegalese population is expected to reach 40 million by 2050 (FAOSTAT, 2019). Pearl millet significantly helps mitigate malnutrition, as it is more nutritious than maize, rice, wheat, and sorghum because its grain contains higher levels of protein, vitamins, amino acids, antioxidants, fibers, and essential

micronutrients (Saleh et al., 2013; Serba et al., 2017). Additionally, this crop has a high potential for transformation into value-added products. Pearl millet grain is used to prepare many traditional foods and is consumed in both rural and urban areas in many forms, such as porridges, gruels, doughs, couscous, and breads. It is also used for forage purposes.

All actors in the Senegalese pearl millet value chain recognized the need to develop strategies to substantially enhance grain yield. Therefore, the breeding program is under time pressure to act as a bridge between the market demand and the crop productivity and should rapidly adopt an efficient breeding strategy to accelerate the improvement of new varieties with better resilience to climate variability in particular drought stress. This chapter highlights the progress in pearl millet breeding in the last several decades, identifies research gaps, and formulates actionable future strategies for improvement of this nationally important cereal crop, particularly in the context of climate change.

1.3 Status and Constraints for Pearl Millet Cultivation in Senegal

Pearl millet is the main staple cereal crop in Senegal, with approximately 800,000 ha of growing area, occupying about 60% of the crop area harvested (FAOSTAT, 2019). It is mainly cultivated in the central, southern, and south-eastern parts of the country, more specifically in the Fatick, Kaffrine, Kaolack, Diourbel, Thies, Tambacounda, and Kolda regions (Figure 1). In 2019, its production in these regions represented more than 80% of total national production (ANSD, 2020). As the populations living in these regions depend particularly on pearl millet production for their livelihoods, improved production technologies are highly sought.

The total national production of pearl millet was 574,000 tons, which accounted for at least 33% of the total cereal production, behind only rice (44%) in 2018 (FAOSTAT, 2019). These ratios among different crops change over years depending on the financial support received for different grain crops (Figure 2A). Pearl millet has occupied the largest area harvested among the cereal crops, followed by rice, sorghum, and maize over the years (Figure 2B). However, the average grain yield remains very low and unstable, never reaching the threshold of 1000 kg ha^{-1} (Figure 2C). By comparison, pearl millet yields in India are 1,243 kg ha^{-1} for the same year (FAOSTAT, 2019). The adoption of hybrids has led to a significant improvement in productivity in India. The first public system for robust hybrid selection of pearl millet was established in the 1960s. This system has also been supported by the private sector and has facilitated significant improvement of productivity in the country with an average increase in grain yield of 24 kg/ha/year between 1996 and 2012 (Yadav & Rai, 2013). In 2006, over 70 hybrid varieties

covered 60% of the cultivated area of pearl millet in India and these hybrid varieties were mainly developed by the private sector and consisted of more than 80% of varieties used at that time (Karandikar et al., 2018). In contrast, no hybrid variety was cultivated in Senegal.

In addition to the absence of hybrid varieties in cultivated areas, growing unimproved varieties has been the major constraint of pearl millet grain productivity (Kanfany et al., 2020). In 2009, only a few improved varieties were cultivated in 34.5% of the national production area. The varieties Thialack 2, Souna 3, and IBMV 8402 covered only 16.5%, 14%, and 4% of the area, respectively (Walker, 2015). In addition, various biotic and abiotic stress conditions, including *Striga* weed (*Striga hermonthica)*, low soil fertility, drought, and insect pests are limiting factors for yield. Moreover, lack of machinery for sowing, disease and bird damage, limited access to certified seeds of high-yielding varieties, and cultivable land all affect production of pearl millet in Senegal (Kanfany et al., 2020). Meanwhile, significant improvement has been made for rice and maize, with an increased productivity of 66% and 282%, respectively, mainly due to the use of higher-yielding varieties and pertinent crop management practices to alleviate biotic and abiotic stresses.

In recent years, the reduction in rainfall volume and the short period of rainfall during the growing seasons have negatively impacted the performance of pearl millet in Senegal (ANSD, 2020). As farmers usually plant pearl millet while waiting for or in response to the first rain of the year, the crop can face severe drought stress in early stages. It is proven that drought stress at an early stage significantly reduces the plant height, biomass, and total grain number and weight in pearl millet inbred lines (Debieu et al., 2018). Early ending of the rainy season can also lead to terminal drought episodes at the grain filling stage, especially for late- maturing varieties. According to the current climate models used in Saharan Africa, including Senegal, more inter- and intra-annual variability in the rainfall amount and distribution is expected (Brown & Lall, 2006), increasingly threatening pearl millet productivity in the country. Several studies have predicted that changed climatic conditions in Sub-Saharan Africa will lead to significant and increasing yield losses for major crops by the 2050s if cropping systems do not adapt. The associated yield losses are forecast at 17% for millet (Schlenker & Lobell, 2010). The Groundnut Basin agro-ecological zone is the most important part of the national pearl millet production area, but it is characterized by a Sudano-Sahelian climate and will be particularly affected by this climate variability.

1.4 Past Attempts of Genetic Improvement and Achievements

1.4.1 Phase 1: Exploitation of the Local Germplasm for Grain Yield Improvement

The pearl millet improvement program started in Senegal in 1931 at the National Center for Agronomic Research (CNRA) in Bambey. There are two ecotypes of pearl millet traditionally grown in Senegal based on life cycle length: the short cycle ecotype (65–90 days) called 'Souna' and the longer cycle (120–150 days) called 'Sanio'. Souna morphotypes flower from 50 to 60 days after planting and are adapted to ecosystems with low and irregular water supplies (350–600 mm annually rainfall). They represent 85% of total millet production and cover the most important part of the nationwide production area in the north and central regions (Diack et al., 2017). Sanio morphotypes flower from 80 to 110 days after planting and are mainly cultivated in the southern region, receiving more important rainfall (900–1200 mm).

During the first phase of the breeding program, the selection was mainly focused on producing inbred lines using pedigree selection from the two ecotypes that were collected locally across the country (Fofana, 1987). This approach led to the development of 113 inbred lines by 1949. However, due to the cross-pollinated nature of pearl millet, these lines lost vigor (Jones & Mangelsdorf, 1925). For instance, one of the local populations gave 20% more grain yield when compared to the performance of the inbred lines selected from this population (Fofana, 1987). Bono & Leclercq, 1963 realized that this breeding method could not be directly used since it only allowed the development of interesting, inbred lines. However, when the inbred lines were crossed with local populations, the top-cross hybrids produced up to 147% yield compared to the local population. By 1959, 28 inbred lines were selected and evaluated for overall performance in three consecutive years based on uniformity, yield performance, and yield component traits. The results showed that even if the homogeneity was significantly improved, the grain yield of these lines was not better than farmers' cultivars (Fofana, 1987). Thus, pedigree selection breeding was judged as an unsuitable tool for pearl millet yield improvement.

In 1961, the national breeding program shifted from pedigree selection to recurrent selection using five local populations: three early (Souna) and two late (Sanio) pearl millet ecotypes (Etasse, 1965). The varieties, derived from the late populations, showed low performance in the first generation of recombination and selection was focused on the early maturing populations. The best progeny was selected from the three early populations and combined to create a synthetic variety named Souna 2. This variety showed high grain yield potential with a 107% and 121% yield increase compared to landraces in the first and second generation of

recombination, respectively. However, the performance in farmer's fields was still variable, mostly lower than local varieties in farmers' fields (less than 2 t/ha). This variety was further improved by top-crossing with a local population (PC28)s using Souna 2 as the tester, resulting in the selection of eight elite lines from this population. The eight lines were combined in 1969 to form Souna 3, the first variety released by the breeding program (MAER, 2012). This synthetic variety was proposed for cultivation in the central and southeastern part of the Groundnut Basin agro-ecological zone (Table 1). This variety has a compact and cylindrical panicle, a resistance to smut and downy mildew, and a grain yield potential of 2.5 to 3.5 t ha-^{1}. It has since been one of the most cultivated varieties in Senegal.

1.4.2 Phase 2: Genetic Improvement of Harvest Index and Resistance to Drought

To achieve self-sufficiency through the improvement of harvest index in local populations, a multi-disciplinary pearl millet program involving several scientists was launched in 1970 (Fofana, 1987). This was one of the first initiatives to strengthen the breeding program after independence in 1960. Thus, 29 populations that were collected from the five main West African countries growing pearl millet were crossed with three dwarf parents sourced from the USA [23 $D_2$5 and 239 D_2B and India (I 472)] (Fofana, 1987). This program developed two dwarf synthetic varieties, GAM 73 and GAM 75 (Table 1) from I 472 x HK 1133 and Tift 23 D2 x Aniata, respectively (Bilquez, 1975). These varieties performed well under irrigated and highly fertilized conditions but could not express their potential under regular farmers' field conditions.

From the 1970s, Senegal experienced a period of severe drought that occurred throughout the entire Sahelian region, affecting crop production considerably (Nouaceur et al., 2017). In addition, the improved varieties in Senegal, as in most West African countries, were not widely diffused throughout the production area due to large environmental variations (Spillane & Gepts, 2001). The multidisciplinary team also developed early maturing (60 and 75 days) synthetic varieties, which were evaluated for overall performance from 1977 to 1979. However, these early varieties produced lower grain yield than Souna 3, even in the environments for which they were developed.

Table 1: Varieties Released for Pearl Millet Cultivation in Senegal

Variety Name	Type	Year of Creation	Year of Release	Maturity Cycle (days)	Potential Grain Yield (t ha-^{1})
Souna 3	Synthetic	1969	1972	90	3.5

IBV 8001	Synthetic	1980	1987	80	3.4
IBV 8004	Synthetic	1980	1987	75	2.6
IBMV 8402	Synthetic	1984	1987	80	2.5
ISMI 9507	Synthetic	1995	2010	85	2.5
Gawane	Composite	2006	2010	85	2.5
Thialack 2	Composite	2008	2010	95	3
Souna Sine	Composite	2018	2021	90	3
Souna Saloum	Composite	2018	2021	90	3
Souna Baol	Composite	2018	2021	90	2.5
Taaw	Hybrid	2017	2021	75	3.5

In 1974, the Senegalese Institute of Agricultural Research (ISRA) was established and started a new millet breeding program with the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in 1977. This millet improvement program was funded by the United Nations Development Program (UNDP) with the main goal of diversifying the genetic bases to develop new varieties that combine stress tolerance and high grain yield potential. This effort made it possible to introduce several millet accessions from India. However, these accessions were not adapted to Senegalese growing conditions because they matured too early,susceptible to diseases and insects, had small panicle size, and produced a similar grain yield as Souna 3. Besides the fact that the size of the panicle has been correlated with the grain yield, the length of the panicle is an important trait for the farmers related to the local practice during harvest. Indeed, the panicles are grouped and tied in heaps after harvest to facilitate transport.

Therefore, these accessions were crossed with local populations to generate new breeding materials, leading to the release of three varieties in 1987 (IBV 8001, IBV 8004, and IBMV 8402) to the central and northern part of the Groundnut Basin agro-ecological zone (Table 1). These varieties were early-maturing and tolerant to downy mildew (Gupta, 1986; Gupta & Ndoye, 1991). In order to boost the national pearl millet production, a small hybrid breeding program was initiated in 1982. Thus, several crosses were made between the selected inbred lines and the landraces with CMS lines introduced from ICRISAT in India. The best performing hybrid (ICMH 8413 from a cross of 81A x IBMI 8207) produced up to 50% more grain yield than Souna 3 (Ndoye & Gupta, 1987). However, these hybrids were not widely grown due to their downy mildew susceptibility, short panicle size, and dwarf stature as farmers preferred tall cultivars for complementary forage yield.

In the early 2000s, the West African National Agricultural Research Systems (NARS), ICRISAT, coordinated with Réseau Ouest et Centre Africain de Recherche sur le Mil

(ROCAFREMI) and International Sorghum and Millet Collaborative Research Support Program (INTSORMIL) regional networks to characterize millet accessions collected from West and Central Africa (Haussmann et al., 2006). Selected accessions were evaluated in regional trials to identify specific varieties that performed well within a country and/or across different agro-ecological zones in West Africa (Wilson et al., 2008). In Senegal, ICTP-8203 and GB-8735 were identified for the Sahelian zone, ISMI-9301 and ISMI-9305 for Sudano-Sahelian zone, and SOSAT-C88 and ICMV-IS-88305 for the Sudanian zone. These varieties increased grain yield up to 22%, on average, compared to the previously released varieties.

1.4.3 Phase 3: Current Genetic Improvement of Pearl Millet in Senegal: A Decade of Rapid Progress

Since 2007, millet genetic improvement in the National Pearl Millet Program has been financially supported by the West Africa Agricultural Productivity Program (WAAPP), a 10-year program funded by the World Bank and several other international funding programs. These funds offered new opportunities to the national program. Significant efforts have been made by the millet program to assemble and define a national core collection, improve local populations, and train young scientists. Progress has also been made on strengthening collaboration with National Agricultural Research Institutes and Advanced Research Centers for Germplasm Exchange and Evaluation, integrating new tools, such as molecular markers and high-throughput phenotyping with conventional breeding, and defining breeding product profile.

Genetic diversity of the germplasm used in a breeding program determines the potential genetic gain achievable through selection (Ambati & Singh, 2017). The national breeding program has made a particular effort to understand, conserve, exploit, and manage the available genetic resources across the country. The national pearl millet collection was updated in 2014 for better representation of the different varieties covering the entire national territory, especially the late flowering varieties (Diack et al., 2017). This new collection contained 477 accessions, including 353 early flowering (Souna), 112 late flowering (Sanio) and 12 improved varieties. These accessions showed high genetic diversity (He = 0.55). This consistently resulted in a greater genetic diversity in Senegalese landraces compared to other regions of Africa and Asia and was surely linked to the proximity of the center of domestication of the crop (Hu et al., 2015).

To facilitate the exploitation of all this diversity in the breeding program, a core collection was established based on phenotypic and genotypic data (SSR markers), using an advanced maximization strategy with a heuristic approach (Kim et al., 2007). This core collection consisted

of 91 accessions (22% of total accessions), including 31 Sanio and 60 Souna, that captured as much genetic diversity as possible from the collected genetic material. The assessment of this core collection's structure revealed two genetic pools (genetic differentiation between Sanio and Souna). Each pool was further divided into three phenotypic subgroups based on the targeted phenotypic traits, such as yield and flowering time. Phenotyping of this core collection revealed five traits, including two phenological traits (heading and flowering) and three morphological traits (biomass, plant height, and tillering) that could clearly differentiate the Souna ecotype from the Sanio ecotype (Diack et al., 2020).

These identified groups and subgroups constituted potential heterotic groups that could be exploited in the millet breeding program to improve the phenotypic traits associated with these differentiations. Heterotic groups are very important for hybrid breeding because they guide the choice of parental lines, optimize the use of genetic material, and maximize the performance of derived hybrids (Meena et al., 2017). The Senegalese breeding program recently initiated studies to identify and validate heterotic groups in the core collection for hybrid breeding as part of a USAID-funded project (Genetic Enhancement of Pearl Millet). This study identified the most representative genotypes of the core collection and produced F1s from crosses between these individuals. Evaluation of these F1s and their parents for general and specific combining ability is currently in progress.

As performance of hybrid lines can be predicted based on general combining ability, trait introgressions in recurrent selection schemes are more efficient using divergent heterotic groups than genetically similar heterotic groups (Reif et al., 2007). Evaluation of hybrids, generated by crossing 17 open-pollinated varieties (OPVs) in a diallel mating design in nine environments over two years in Niger and Senegal resulted in the identification of 5 potential parental lines (Nigerien CIVT, H80-10Gr, Taram and Senegalese Thialack 2, and Souna 3). These five parental lines have a high combining ability that can be used to initiate reciprocal recurrent selection (Sattler et al., 2019).

The combining ability and heterosis for grain yield and its associated traits (i.e., plant height, flowering time, panicle length and diameter, productive tillers, thousand grains weight, and panicle and downy mildew) were assessed by crossing 17 inbred lines with Sosat C88 and Souna 3. The results indicated that the inheritance of these traits was controlled by both additive and nonadditive gene effects although the contribution of the additive genes was greater (Kanfany et al., 2018a). Some inbred lines (IBL003-B-1, IBL091-1-1, IBL095-4-1, IBL110-B-1, and IBL206-

1-1) showed positive General Combining Ability (GCA) effects on grain yield and negative GCA effects on downy mildew, flowering time, and plant height; therefore, they are good parents for breeding high-yielding open-pollinated varieties and hybrids (Kanfany et al., 2018b).

Crop breeding is now undergoing a paradigm shift with the development of modern approaches. These methods, such as marker assisted selection or genomic selection, are more accurate and less costly in time and effort and are a byproduct of next-generation sequencing (NGS). In an effort to understand the genomic diversity and population structure of Senegalese germplasm, all accessions in the core collection were characterized using Genotype By Sequencing (GBS). This yielded 21,663 polymorphic (biallelic) and high-quality Single Nucleotide Polymorphism (SNP) markers. Marker-trait association identified markers linked to the adaptive genes *PgPHYC* and *PgMADS11* that were involved in the variation of flowering time in millet (Diack et al., 2017). Eighteen SNPs and genes have been associated with flowering time, plant height, number of tillers, and biomass, with some of these genes acting as candidate genes for abiotic stress adaptation. These identified markers, after validation, will be useful for marker-assisted selection in the program.

Cost and time efficiency in breeding programs depend on the ability to identify parental lines at an early stage of the breeding scheme (Ambati & Singh, 2017). To this end, studies were carried out to identify donor parents for the traits of farmers' interest. Firstly, genotypes with desirable agronomic traits and stable resistance to downy mildew (DM) were identified in hotspot areas (Zoclanclounon et al., 2019). Then, inbred lines, derived from a collection of landraces originating from West and Central African countries, were evaluated in the Bambey and Nioro Research Stations under field conditions with spreader rows of DM, using SOSAT-C88 and 7042S as resistant and susceptible controls, respectively (Kanfany et al., 2018b). Highly significant differences were observed among 101 lines tested for DM incidence and severity as well as agro-morphological traits, including plant height, flowering time, panicle length, and productive tillers. Based on DM incidence and severity of control varieties used, the testing lines were grouped into three categories: resistant (55, including SOSAT-C88), moderately resistant (16), and susceptible (30, including 7042S). Also, using DM resistance and plant height as the most discriminant factors, hierarchical ascendant cluster analysis grouped these entries into three clusters, confirming the previous categorization.

Conversely, the released Senegalese varieties (Gawane, ISMI9507, Thialack2 varieties, IBV8004, and Souna 3) were evaluated for resistance to stem borers and maize head miner (MHM)

under natural infestation in the 2014 and 2015 cropping seasons (Goudiaby et al., 2018). Gawane was identified to be tolerant to stem borers and Thialack2 and ISMI9507 were identified to be resistant to MHM (Goudiaby et al., 2018). This effort to identify new sources of resistance to several diseases under Senegalese agro-ecological conditions provides new genetic resources for improvement of pearl millet for agronomic traits and resistance to DM, MHM, and stem borer. These new resources will be used in the national breeding program to combine into new varieties resistant to biotic stresses and with desirable adaptation traits.

All these efforts allowed the breeding program to select and propose four new pearl millet varieties for release in 2021, including three OPV (SL28, SL169, and SL423) and one top-cross hybrid. The OPVs were derived from local populations, which were improved through recurrent selection using the S1 method. The top-cross hybrid named TAAW was created by crossing a local population (SL175) to a male sterile parent sourced from ICRISAT (ICMA-177090).

1.5 *Productivity Improvement Over the Different Phases*

The average rates of yield gains from each of the three phases of pearl millet improvement in Senegal are listed in Figure 3. In the first phase (1961–1973), low average grain yield was observed throughout the country, mainly due to drought stress in early stages and lack of adapted varieties (ANSD, 2020). Thus, the breeding efforts at that period were focused on the development of varieties adapted to specific agroclimatic zones. The breeding effort in the second phase (1974–2007) led to the release of several new varieties, with an increased yield of 2.06 kg/ha. The highest rate of productivity was achieved in the third phase (2008–2019), with a yield gain of 2.7 kg/ha/yr. However, the genetic gain through these three phases remains unknown.

Assuming that this continued increase in average grain yield has been achieved through genetic gain, it promises great potential for further improvement in millet yield, especially with the help of newly developed genomic tools. However, it should be considered that this rapid increase in grain yield may be caused by the combination and fixation of favorable alleles at major effect loci that underlie traits positively correlated with grain yield. Indeed, selection has been shown to generally fix favorable alleles at major effect loci faster than those at minor effect loci (Mayo & Hancock, 1981). In this case, it would be expected that the yield improvement be less significant in the coming years since it would be mainly driven by the fixation of favorable alleles at loci with minor effects and would depend more heavily on better management practices (Mayo & Hancock, 1981; Xu et al., 2017; Yeaman, 2022). Nevertheless, the breeding history described above and the

current breeding pipeline which regularly brings genetic material from the associated programs suggest that the large effect variants have not yet segregated into the breeding germplasm (Mayo & Hancock, 1981; Muleta et al., 2019). Therefore, more significant yield improvement through genetic improvement should reasonably be targeted.

1.6 Future Genetic Improvement Strategies

Grain yield is unanimously considered the most important trait by farmers (Kanfany et al., 2020). Drought tolerance, adaptation to low soil fertility, grain color, disease and insect resistance, and early maturity are also important traits to be considered when farmers select a pearl millet variety to grow in their fields (Kanfany et al., 2020). Therefore, these should be the target traits for developing a variety that grows in Senegalese conditions. However, the order of priority for these traits varies from location to location. Incorporating farmers' preferred basic and value-added traits into the breeding product profile would certainly increase farmer acceptance of released varieties.

The breeding goal in Senegal is to release pearl millet varieties with high yields (15% increase in grain yield over Souna 3 and Thialack II), high iron contents (with > 40 ppm Fe), and better adaptation in the two agro-ecological zones (Table S1). The combined strategies described below should enable the National Pearl Millet Improvement Program to diversify its sources of germplasm and accelerate varietal development with better selection accuracy.

1.6.1 Towards Implementing Genomics-enabled Breeding

In the current breeding scheme, phenotypic selection remains the major method used in Senegal. However, this approach may not be able to properly examine the hidden genetic variation available for complex agronomic traits, such as grain yield and drought resistance, as these traits typically have low heritability and are heavily influenced by environments (Wu et al., 2012). Moreover, phenotypic selection involves many selection cycles in multi-trials in different environments to accurately evaluate progeny performance, which not only needs large amounts of seed for replicated trials, but is also laborious, time-consuming, and costly. However, such tests may not be realized due to insufficient seed amount and financial support for resource-limited breeding programs of developing countries like Senegal. Furthermore, the current and predicted variation in rainfall amount and distribution in the Sahel zone has emphasized the difficulty to control genotype x environment (GxE) interactions.

In light of the challenge of increasing genetic gain with limited resources, the National Pearl Millet Improvement Program has considered making greater use of currently available genomic resources to help breeders make more informed decisions about crosses at early breeding stages. The availability of the pearl millet reference genome sequence (~1.79 Gb) and the reduction in the cost of high-throughput data acquisition offers the opportunity to strengthen the breeding pipeline (Varshney et al., 2017). Nevertheless, the choice of modern breeding tools and methods to integrate must be evaluated with the utmost caution, considering several factors, including the genetic architecture of the targeted traits. For instance, the results of simulations performed by Muleta et al. (2019) showed that there are specific conditions under which genomics-assisted recurrent selection (GARS) should be considered by small breeding programs, according to the genetic architecture of the trait. In fact, this study showed a lower cost per unit gain using phenotypic recurrent selection (PRS) compared to GARS for oligogenic traits in small new breeding programs, while GARS gave lower cost per unit gain than PRS with larger populations and polygenic traits. Thus, understanding the genetic architecture of target traits will guide the breeding program in adopting the most effective breeding pipeline to maximize genetic gain given available resources.

Genomics-assisted breeding that integrates advanced genomic tools with conventional breeding has been one of the main strategies for developing high-yielding crop varieties that are resistant to biotic and abiotic stresses. Over the past decade, the National Pearl Millet Improvement Program has focused on understanding the genomic diversity of available germplasm, identifying genetic variants of targeted traits, and uncovering the molecular and physiological mechanisms underlying the expression of these traits. Phenotypic variants identified for downy mildew, MHM, and drought tolerance and allelic variants (SNPs) identified for flowering time provide new genetic resources for the development of molecular markers linked to favorable alleles. Therefore, more accurate selection for these traits could be achieved using marker-assisted backcrossing to shorten breeding cycles (Bernardo, 2016). In the years to come, the National Pearl Millet Improvement Program will take advantage of new financial resources from the Sorghum and Millet Innovation Lab (SMIL), Innovation Lab for Crop Improvement (ILCI), and Desira-ABEE to set up a strong marker-assisted breeding program to accelerate the introgression of these genetic variants into elite lines to improve the genetic gain.

1.6.2 Breeding for Drought Adaptive Traits in the Senegalese Agro-ecosystem

Drought tolerance is a polygenic trait significantly influenced by the environment (Serba & Yadav, 2016). The development of drought-tolerant varieties is a challenging task that can be achieved through the introgression of adaptive traits that allow plants to grow and produce satisfactory yields under water-limited conditions. In some cases, complex traits, such as drought, can be more efficiently improved by targeting simpler and more strongly correlated traits with them (Breseghello & Coelho, 2013). Drought adaptive traits can be categorized into three main strategies: drought escape, dehydration avoidance, and dehydration tolerance (Basu et al., 2016). Drought escape allows plants to accelerate their life cycle and reproduce before drought occurs (Turner, 1896). This strategy is commonly used by annual plants although the yield is generally decreased (Lisar et al., 2016). Dehydration avoidance is a strategy for preserving the high-water potential of plants mainly by reducing water loss through transpiration and optimizing water uptake from the soil by the root system (Basu et al., 2016). Because the mechanisms involved in this strategy consume additional energy, plant yield is generally decreased and the plants are small (Lisar et al., 2016). Avoidance is mainly achieved through traits related to root architecture, stomatal control, and transpiration efficiency (Basu et al., 2016). Dehydration tolerance involves mechanisms to maintain functionality in a water stressed environment (Turner, 1896). Dehydration tolerance involves traits such as cell remobilization of stem water-soluble carbohydrates, membrane stability and osmotic adjustment (Lisar et al., 2016). A drought tolerance cultivar can be defined as a genotype that detects water deficiency and effectively adopts combined strategies to grow and produce a satisfactory yield. Understanding the traditional cropping systems' response to water deficiency and identifying the valuable traits to be targeted are critical steps in breeding a drought tolerant cultivar.

Through evolution, traditional pearl millet landraces showed evidence of different mechanisms of adaptation to climate change, resulting from natural selection as well as from farmers' selection. On the one hand, the development of the root system of pearl millet is distinguished by a fast-growing primary root that rapidly colonizes the deeper horizons of the soil at an early stage (Passot et al., 2016). This trait allows access to the water available in a deeper soil horizon. Comparison of root development parameters in 16 inbred lines originating from Indian, West African, and Central African landraces showed significant variation in primary root length and lateral root density that could be exploited in subsequent breeding efforts (Passot et al., 2016). On the other hand, West African pearl millet farmers preferably cultivate high tillering landraces

with small grains (Serba & Yadav, 2016). These landraces can produce secondary tillers and fill the grain very quickly. Since yield potential is related to the number of tillers and grains, among other morphological traits, this strategy allows them to efficiently improve yield under drought conditions. Finally, evidence is beginning to accumulate that flowering time is under selection in West Africa in response to rainfall variability. Flowering time directly correlates with annual rainfall (Saïdou et al., 2009).

Comparison analysis of pearl millet accessions collected in 1976 and 2003 in Niger showed a significant shift: varieties collected in 2003 flowered slightly earlier and had a shorter spike than in 1976 (Saïdou et al, 2009). This response to selection for flowering time was significantly associated with polymorphisms in two genes *PHYC* and *PgMADS11* (Mariac et al., 2011; Saïdou et al., 2009). These polymorphisms for early and late flowering alleles in these two genes were found in the landraces in Niger and were also more recently revealed in the Senegalese germplasm (Diack et al., 2020; Vigouroux et al., 2011). These findings suggest that drought escape is a main mechanism of adaptation to drought in West Africa. Understanding the genetic basis of these morpho-physiological traits associated with drought tolerance, in particular the flowering period, and then identifying the favorable alleles to introgress into the elite background are promising steps for the rapid selection of a drought-tolerant cultivar.

1.6.3 Reinforcement of Hybrid Breeding Strategies

In view of the urgent need for a rapid increase in productivity, it is crucial to establish a strong hybrid breeding program to focus on breeding high-yielding hybrid cultivars. Combining ability studies have clearly highlighted the prevalence of high levels of heterosis and potential to breed for hybrids with high and stable yield, as well as multiple disease resistance and abiotic stress tolerance in Senegalese environments. Previous studies showed the superiority of hybrid varieties over the best landraces for agro-morphological and agronomic traits in West African pearl millet germplasm (Pucher et al., 2016).

Inspired by the success of hybrids in India, a collaboration has been initiated by millet breeding programs in several West African countries (Senegal, Niger, Burkina Faso, and Mali) as part of the SMIL project for the development of dual-purpose pearl millet hybrid lines with enhanced grain nutritional quality and stover digestibility. The aim of the SMIL project was to characterize genetic diversity, validate superior germplasm accessions, and identify genetic variants for traits related to stover quality and digestibility (i.e., grain mineral content, grain and stover yields) using 100 germplasm accessions collected from the related countries (Serba et al.,

2020). Adoption of dual-purpose pearl millet hybrid varieties are expected to improve nutritional security of smallholder farming families, crop-livestock integration, and income generation.

1.7 Conclusion

Pearl millet is the main cereal grown in Senegal. The traditional farming system has played a crucial role in creating and maintaining a high diversity among landraces. The Senegalese breeding program has invested a great deal of effort for nearly a century in collecting, maintaining, and more recently understanding this genetic diversity. In parallel, several varieties were derived from the local population using the conventional breeding approach. The regional and international partnerships developed have made it possible to enrich the genetic material and support all these selection efforts, resulting in a significant improvement in yield performance, especially in the last phase ranging from 2008–2019. This last decade was particularly marked by progress towards the implementation of genomics-based breeding and the strengthening of hybrid breeding strategies. Despite their yield performance, the varieties released so far do not seem to hold all the agro-morphological characteristics that lead to their high adoption by farmers. The high genetic variation found in Senegalese landraces that possess adaptive traits to the local environment, as well as traits preferred by farmers, is a promising resource for developing high-yielding crop varieties that are resistant to specific biotic and abiotic stresses. In addition, the future breeding strategies outlined in this chapter, and the more participatory breeding approach should enable the breeding program to improve genetic gain more effectively for farmers' priority traits. Further, they will contribute to addressing the challenge of sustainably improving pearl millet productivity for food security and income generation in Senegal.

Acknowledgements

This study is made possible by the support of the American People provided to the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-A13-00047. The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

Chapter 1 References

Ambati, S. & Singh, T. V. J. (2017). Heterotic grouping and its importance in tropical hybrid rice breeding. *International Journal of Multidisciplinary Advanced Research Trends*, *9*(1), 179–187.

ANSD. (2018). Sénégal : Enquête démographique et de santé continue (EDS-Continue) 2017. Agence Nationale de la Statistique et de la Démographie (ANSD).

ANSD. (2020). Situation économique et sociale du Sénégal. Agence Nationale de la Statistique et de la Démographie (ANSD).

Basu, S., Ramegowda, V., Kumar, A., & Pereira, A. (2016). Plant adaptation to drought stress. F1000 Research 5, F1000 Faculty Rev-1554. https://doi.org/10.12688/f1000research.7678.1

Bernardo, R. (2016). Genomewide predictions for backcrossing a quantitative trait from an xxotic to an adapted line. *Crop Science*, *56*, 1067–1075. https://doi.org/10.2135/cropsci2015.09.0586

Bilquez, A. F. (1975). Amélioration des mils au Sénégal: synthèse des résultats obtenus au cours des quatre premières années de travail et conclusions générales. I.S.R.A., 57 p. multigr.

Bono, M., & Leclercq, P. (1963). Méthodes d'amélioration variétale des mils et sorghos utilisées au CRA Bambey. *L'Agronomie Tropicale. Série 1, Riz et Riziculture et Cultures Vivrières Tropicales, 18*(1), 33–52. https://agritrop.cirad.fr/444782/

Breseghello, F., & Coelho, A. S. G. (2013). Traditional and modern plant breeding methods with examples in rice (Oryza sativa L.). *Journal of agricultural and food chemistry, 61*(35), 8277–8286.

Brown, C., & Lall, U. (2006). Water and economic development: The role of variability and a framework for resilience. *Natural Resources Forum, 30*(4), 306–317. https://doi.org/10.1111/j.1477-8947.2006.00118.x

Burgarella, C., Cubry, P., Kane, N. A., Varshney, R. K., Mariac, C., Liu, X., Shi, C., Thudi, M., Courdec, M., Xu, X., Chitikineni, A., Scarcelli, N., Barnaud, A., Rhoné, B., Dupuy, C., François, O., Berthouly-Salazar, C., & Vigouroux, Y. (2018). A western Sahara centre of domestication inferred from pearl millet genomes. *Nature ecology & evolution*, *2*(9), 1377-1380.https://doi.org/10.1038/s41559-018-0643-y

Debieu, M., Sine, B., Passot, S., Grondin, A., Akata, E., Gangashetty, P., Vadez, V., Gantet, P., Foncéka, D., Cournac, L., Hash, C. T., Kane, N. A., Vigouroux, Y., & Laplaze, L. (2018). Response to early drought stress and identification of QTLs controlling biomass production under drought in pearl millet. *PLoS One, 13*(10), https://doi.org/10.1371/journal.pone.0201635

Diack, O., Kane, N. A., Berthouly-Salazar, C., Gueye, M. C., Diop, B. M., Fofana, A., Sy, O., Tall, H., Zekraoui, L., Piquet, M., Couderc, M., Vigouroux, Y., Diouf, D., & Barnaud, A. (2017). New genetic insights into pearl millet diversity as revealed by characterization of early-and late-flowering landraces from Senegal. *Frontiers in plant science*, *8*(818). https://doi.org/10.3389/fpls.2017.00818

Diack, O., Kanfany, G., Gueye, M. C., Sy, O., Fofana, A., Tall, H., Serba, D. D., Zekraoui, L., Berthouly -Salazar, C., Vigourouz, Y., Diouf, D., & Kane, N. A. (2020). GWAS unveils features between early-and late-flowering pearl millets. *BMC Genomics*, *21*(1), 1–11.

Dussert, Y., Snirc, A., & Robert, T. (2015). Inference of domestication history and differentiation between early- and late-flowering varieties in pearl millet. *Molecular Ecology, 24*(7), 1387–1402. https://doi.org/10.1111/mec.13119

Etasse, C. (1965). Amélioration du mil Pennisetum au Sénégal. L'Agronomie Tropicale. Série 1, Riz et Riziculture et Cultures Vivrières Tropicales. https://agritrop.cirad.fr/435118/

FAOSTAT. (2019). Food and Agriculture Organization of the United Nations. FAO Database. Statistiques Agricoles. http://faostat.fao.org.

Fofana, A. (1987). Amélioration du mil au Sénégal. Annual report. Centre Nationale de Recherches Agronomiques. Bambey, Senegal. ISRA, CNRA.

Goudiaby, M. F., Sarr, I., & Sembene, M. (2018). Source of resistance in pearl millet varieties against stem borers and the ear headminer. *Journal of Entomology and Zoology Studies*, *6*(1), 1702–1708.

Gupta, S. (1986). The pearl millet improvement program in Senegal. Annual Report. Centre Nationale de Recherches Agronomiques, Bambey, Senegal. ICRISAT, ISRA, UNDP, Cooperative Programme 1977–1985.

Gupta, S. C., & Ndoye, A. T. (1991). Yield stability analysis of promising pearl millet genotypes in Senegal. *Maydica*, *36*(1), 83–86.

Haussmann, B. I. G., Boubacar, A., Boureima, S. S., & Vigouroux, Y. (2006). Multiplication and preliminary characterization of West and Central African pearl millet landraces. *International Sorghum and Millets Newsletter, 47*, 110–112.

Hu, Z., Mbacké, B., Perumal, R., Guèye, M. C., Sy, O., Bouchet, S., Prasad, P. V. V., & Morris, G. P. (2015). Population genomics of pearl millet (Pennisetum glaucum (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. *BMC Genomics*, *16*(1), 1–12.

Jones, D F, & P. C. Mangelsdorf. (1925). The improvement of naturally cross-pollinated plants by selection in self-fertilized lines: The production of inbred strains of corn. *Bulletin, Connecticut Agricultural Experiment Station*, *266*, 349–418.

Kanfany, G., Diack, O., Kane, N. A., Gangashetty, P. I., Sy, O., Fofana, A., & Cisse, N. (2020). Implications of farmer perceived production constraints and varietal preferences to pearl millet breeding in Senegal. *African Crop Science Journal, 28*(3), 411–420. https://doi.org/10.4314/acsj.v28i3.6

Kanfany, G., Fofana, A., Tongoona, P., Danquah, A., Offei, S., Danquah, E., & Cisse, N. (2018a). Estimates of combining ability and heterosis for yield and its related traits in pearl millet inbred lines under downy mildew prevalent areas of Senegal. *International Journal of Agronomy*. https://doi.org/10.1155/2018/3439090

Kanfany, G., Fofana, A., Tongoona, P., Danquah, A., Offei, S., Danquah, E., & Cisse, N. (2018b). Identification of new sources of resistance for pearl millet downy mildew disease under field conditions. *Plant Genetic Resources*, *16*(4), 397–400.

Karandikar, B., Smale, M., Birol, E., & Tedla-Diressie, M. (2018). India's pearl millet seed industry: Prospects for high-iron hybrids. *Harvest*. Work. Pap., 28, 1–41.

Kim, K. W., Chung, H. K., Cho, G. T., Ma, K. H., Chandrabalan, D., Gwag, J. G., Kim, T. S., Cho, E. G., & Park, Y. J. (2007). PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics, 23*(16), 2155–2162. https://doi.org/10.1093/bioinformatics/btm313

MAER, Ministère de l'Agriculture et de l'Equipement Rural. (2012). Catalogue officiel des espèces et des variétés cultivées au Sénégal. Dakar, Senegal.

Manning, K., Pelling, R., Higham, T., Schwenniger, J. L., & Fuller., D. Q. (2011). 4500-year old domesticated pearl millet (Pennisetum glaucum) from the Tilemsi Valley, Mali: new insights into an alternative cereal domestication Pathway. *Journal of Archaeological Science 38*(2), 312–322. https://doi.org/10.1016/j.jas.2010.09.007

Mariac, C., Jehin, L., Saïdou, A. A., Thuillet, A. C., Couderc, M., Sire, P., Jugdé, H., Adam, H., Bezançon, G., Pham, J. L., & Vigouroux, Y. (2011). Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. *Molecular Ecology*, *20*(1), 80–91. https://doi.org/10.1111/j.1365-294X.2010.04893.x

Mayo, O., & Hancock, T. W. (1981). Fixation of genes having large or small effects on a trait with an intermediate optimum. *Human Heredity*, *31*(5), 286–290. https://doi.org/10.1159/000153224

Meena, A. K., Gurjar, D., Patil, S. S., & Kumhar, B. L. (2017). Concept of heterotic group and its exploitation in hybrid breeding. *International Journal of Current Microbiology and Applied Sciences, 6*(6), 61–73. https://doi.org/10.20546/ijcmas.2017.606.007

Muleta, K. T., Pressoir, G., & Morris, G. P. (2019). Optimizing genomic selection for a sorghum breeding program in Haiti: A simulation study. *G3: Genes, Genomes, Genetics*, *9*(2), 391–401. https://doi.org/10.1534/g3.118.200932

Ndoye, A. T., & Gupta, S. C. (1987). Research on pearl millet hybrids in Senegal in the international crops research institute for the semi-arid. Proceedings of the international pearl millet workshop. ISRA/ICRISAT.

Nouaceur, Z., Murărescu, O., & Murătoreanu, G. (2017). Rainfall variability and trend analysis of multi annual rainfall in Romanian plain. *Annals of Valahia University of Targoviste*, *Geographical Series, 17*(2), 124–144.

Oumar, I., Mariac, C., Pham, J. L., & Vigouroux, Y. (2008). Phylogeny and origin of pearl millet (Pennisetum glaucum [L.] R. Br) as revealed by microsatellite loci. *Theoretical and Applied Genetics*, *117*(4), 489-497.

Passot, S., Gnacko, F., Moukouanga, D., Lucas, M., Guyomarc'h, S., Ortega, B. M., Atkinson, J. A., Belko, M. N., Bennett, M. J., Gantet, P., Wells, D. M., Guédon, Y., Vigouroux, Y., Verdeil, J. L., Muller, B., & Laplaze, L. (2016). Characterization of pearl millet root architecture and anatomy reveals three types of lateral roots. *Frontiers in Plant Science, 7,* 829. https://doi.org/10.3389/fpls.2016.00829

Pucher, A., Sy, O., Sanogo, M. D., Angarawai, I. I., Zangre, R., Ouedraogo, M., Boureima, S., Hash, C. T., & Haussmann, B. I. (2016). Combining ability patterns among West African pearl millet landraces and prospects for pearl millet hybrid breeding. *Field Crops Research, 195,* 9–20. https://doi.org/10.1016/j.fcr.2016.04.035

Reif, J. C., Gumpert, F. M., Fischer, S., & Melchinger, A. E. (2007). Impact of interpopulation divergence on additive and dominance variance in hybrid populations. *Genetics, 176*(3), 1931–1934. https://doi.org/10.1534/genetics.107.074146

Saïdou, A. A., Mariac, C., Luong, V., Pham, J. L., Bezançon, G., & Vigouroux, Y. (2009). Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. *Genetics*, *182*(3), 899–910.

Saleh, A. S. M., Zhang, Q., Chen, J., & Shen, Q. (2013). Millet grains: Nutritional quality, processing, and potential health benefits. *Comprehensive Reviews in Food Science and Food Safety*, *12*(3), 281–295. https://doi.org/10.1111/1541-4337.12012

Salehi-Lisar, S. Y., & Bakhshayeshan-Agdam, H. (2016). Drought stress in plants: causes, consequences, and tolerance. In *Drought Stress Tolerance in Plants, Vol 1* (pp. 1-16). Springer, Cham.

Sattler, F. T., Pucher, A., Kassari Ango, I., Sy, O., Ahmadou, I., Hash, C. T., & Haussmann, B. I. (2019). Identification of combining ability patterns for pearl millet hybrid breeding in West Africa. *Crop Science*, *59*(4), 1590–1603. https://doi.org/10.2135/cropsci2018.12.0727

Schlenker, W., & Lobell, D. B. (2010). Robust negative impacts of climate change on African agriculture. *Environmental Research Letters*, *5*(1).

Sehgal, D., Skot, L., Singh, R., Srivastava, R. K., Das, S. P., Taunk, J., Sharma, P. C., Pal, R., Raj, B., Hash, C. T., & Yadav, R. S. (2015). Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLoS One, 10*(5). https://doi.org/10.1371/journal.pone.0122165

Serba, D. D., Muleta, K. T., St. Amand, P., Bernardo, A., Bai, G., Perumal, R., & Bashir, E. (2019). Genetic diversity, population structure, and linkage disequilibrium of pearl millet. *The Plant Genome*, *12*(3). https://doi.org/10.3835/plantgenome2018.11.0091

Serba, D. D., Perumal, R., Tesso, T. T., & Min, D. (2017). Status of global pearl millet breeding programs and the way forward. *Crop Science*, *57*(6), 2891–2905. https://doi.org/10.2135/cropsci2016.11.0936

Serba, D. D., Sy, O., Sanogo, M. D., Issaka, A., Ouedraogo, M., Ango, I. K., Drabo, I., & Kanfany, G. (2020). Performance of dual-purpose pearl millet genotypes in West Africa: Importance of morphology and phenology. *African Crop Science Journal*, *28*(4), 481–498.

Serba, D. D., & Yadav, R. S. (2016). Genomic tools in pearl millet breeding for drought tolerance: Status and prospects. *Frontiers in Plant Science*, *7*, 1724. https://doi.org/10.3389/fpls.2016.01724

Spillane, C., & Gepts, P. (2001). Broadening the genetic base of crops: An overview. In H. D. Cooper, C. Spillane, & T. Hodgkin (Eds.), *Broadening the genetic of crop production*, (pp. 1–23). CABI Publishing.

Turner, N. C. (1986). Adaptation to water deficits: a changing perspective. *Functional Plant Biology*, *13*(1), 175-190.

Varshney, R. K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., ... & Xu, X. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature biotechnology*, *35*(10), 969-976.

Vigouroux, Y., Mariac, C., De Mita, S., Pham, J. L., Gérard, B., Kapran, I., Sagnard, F., Deu, M., Chantereau, J., Ali, A., Ndjeunga, J., Luong, V., Thuillet, A. C., Saïdou, A. A., & Bezançon, G. (2011). Selection for earlier flowering crop associated with climatic variations in the Sahel. *PloS One*, *6*(5), https://doi.org/10.1371/journal.pone.0019563

Walker, T. S., & Alwang, J. (Eds.). (2015). *Crop improvement, adoption, and impact of improved varieties in food crops in Sub-Saharan Africa*. CABI Publishing.

Wilson, J. P., Sanogo, M. D., Nutsugah, S. K., Angarawai, I., Fofana, A., Traore, H., Ahmadou, I., & Muuka, F. P. (2008). Evaluation of pearl millet for yield and downy mildew resistance across seven countries in Sub-Saharan Africa. *African Journal of Agricultural Research 3*(4), 371–378.

Wu, X., Chang, X., & Jing, R. (2012). Genetic insight into yield-associated traits of wheat grown in multiple rain-fed environments. *PloS One*, *7*(2), https://doi.org/10.1371/journal.pone.0031249

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B., & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. *Journal of Experimental Botany*, *68*(11), 2641-2666.

Yadav, O. P., & Rai, K. N. (2013). Genetic improvement of pearl millet in India. *Agricultural Research, 2*(4), 275–292.

Yeaman, S. (2022). Evolution of polygenic traits under global vs local adaptation. *Genetics*, *220*(1), iyab134.

Zoclanclounon, Y. A. B., Kanfany, G., Kane, A., Fonceka, D., Ehemba, G. L., & Ly, F. (2019). Current status of pearl millet downy mildew prevalence across agroecological zones of Senegal. *The Scientific World Journal*, *2019*. https://doi.org/10.1155/2019/1252653

Chapter 1 Figures and Tables

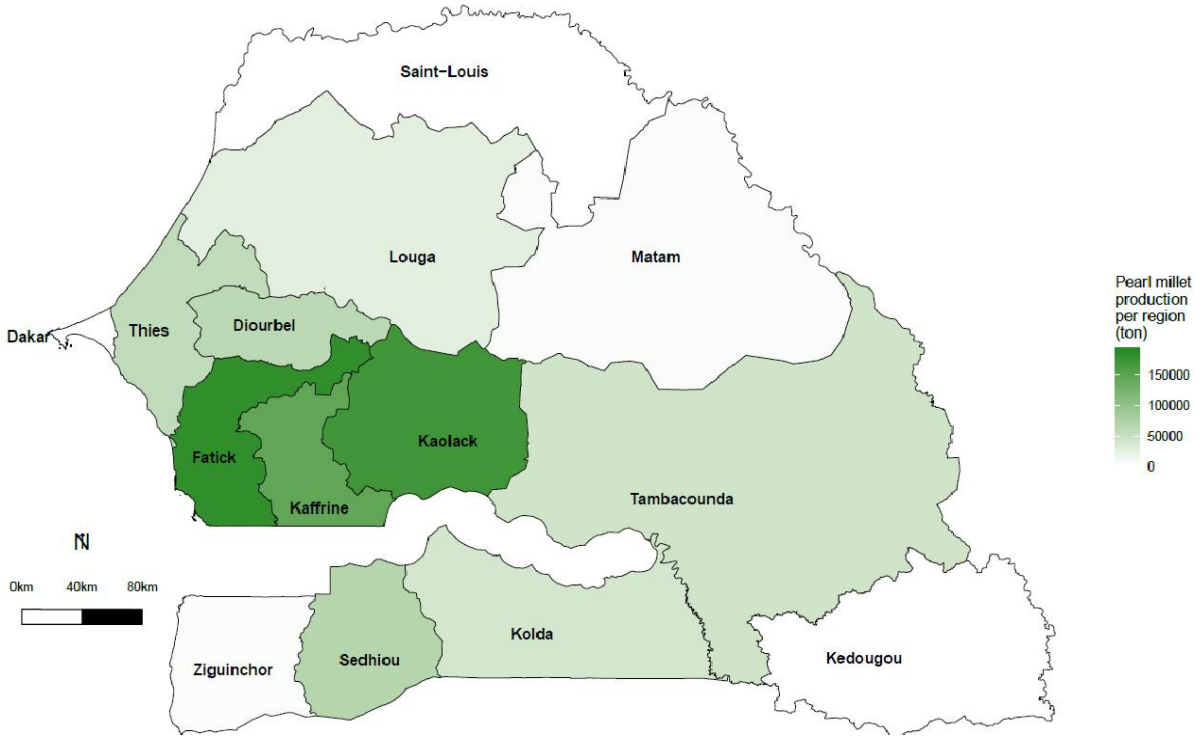
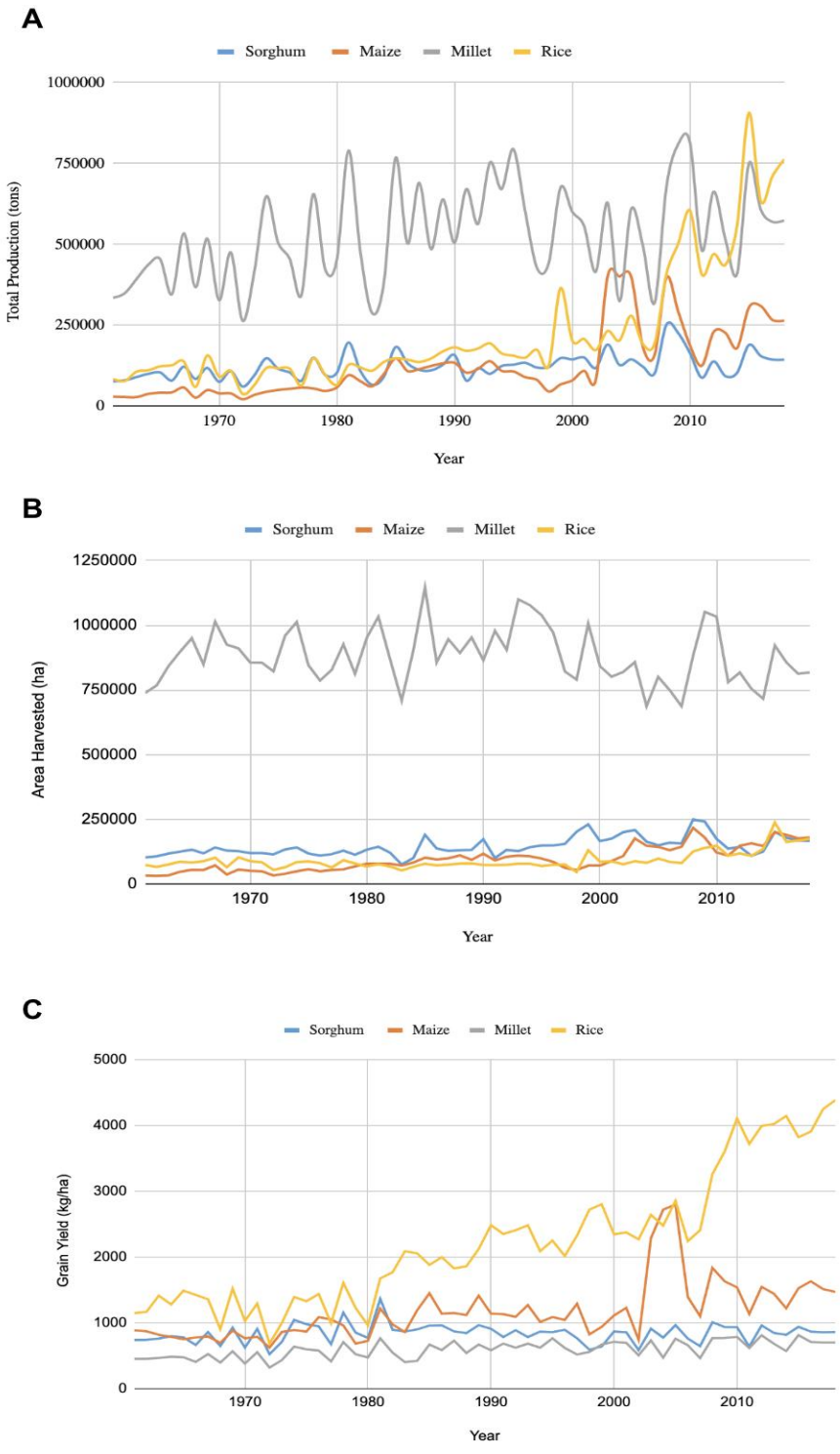


Figure 1. ***Distribution of Pearl Millet Production Areas in Senegal During the Rainy Season of 2020***



Note. From FAOSTAT, 2020.

Figure 2. Dynamics of Cereal Production by Crop in Senegal from 1961 to 2018. (A) Total Production of Cereals (tons); (B) Area Harvested (ha); (C) Grain Yield (kg ha-^{1}).

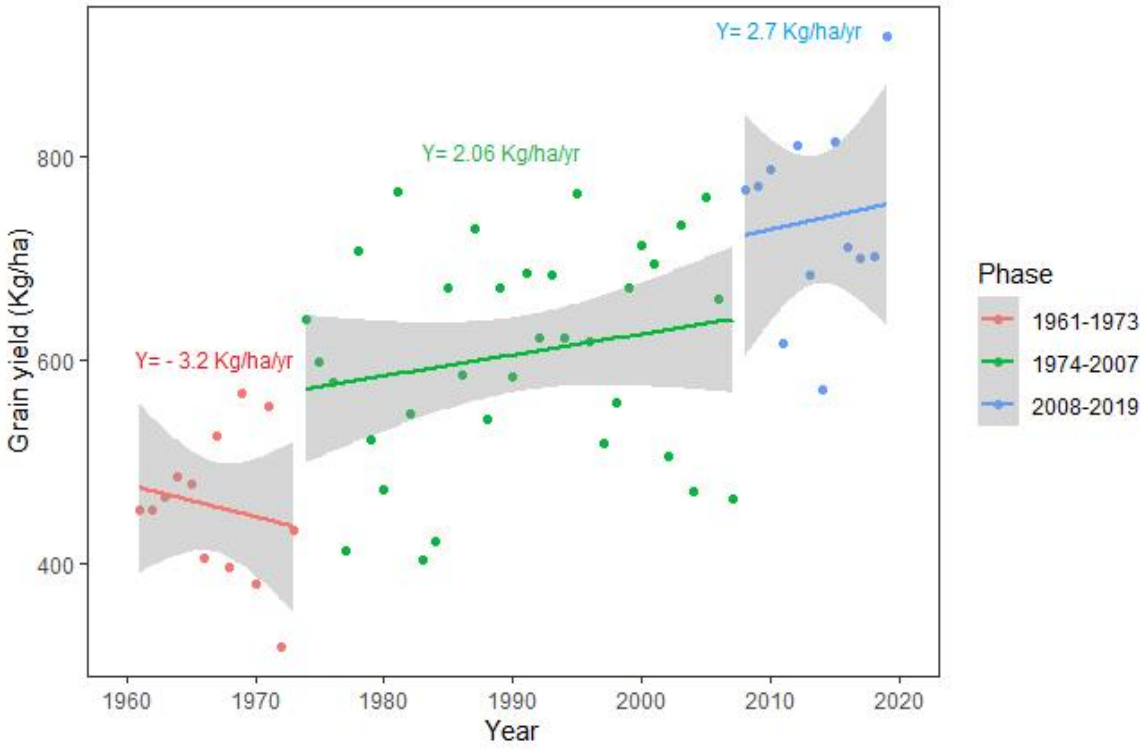


Figure 3. Productivity Improvement Rates for Pearl Millet Productivity During the Three Phases of Pearl Millet Breeding in Senegal: Phase 1 - Exploitation of the Local Germplasm for Grain Yield Improvement (1961–1973), Phase 2 - Genetic Improvement of Harvest Index and Resistance to Drought (1974–2007), and Phase 3 - Current Genetic Improvement (2008–2019). Note. From FAOSTAT, 2019.

Table S1. Breeding targets for the two agro-climatic zones in Senegal where Pearl Millet is mostly produced

Breeding Targets	Target Agro-ecologies	Maturity (days) & Rainfall	Resistance/tolerance Required	Basic Traits	Value Added Traits	Breeding Goals
High yielding, early maturing pearl millet **OPVs** for adaptation to Sahelian zone of Senegal	Target: Central and northern part of the peanut basin agro-ecological zone	Early maturity 70–90 days (350–600 mm/annul)	**Biotic stresses:** Downy mildew, millet head-miner *Striga hermonthica* **Abiotic stresses:** Drought, Low soil fertility	Grain Yield of >2 tons/ha, Downy mildew resistance Head length > 40 cm	High grain iron and zinc content. Striga tolerance Drought tolerance Head miner tolerance	10% increase in grain yield over improved check (Souna 3) with >40 ppm Fe
Medium maturity, high yielding pearl millet **OPV** Soudanian zone of Senegal	Target: Southern part of the peanut basin agro-ecological zone	Medium maturity/ >90 days (500–900 mm/annum)	**Biotic stresses:** Downy mildew, head miner, striga hermonthica **Abiotic stresses:** Drought, low soil fertility	Grain Yield of >2 tons/ha, downy mildew resistance Head length > 50 cm	High grain iron and zinc content. Striga tolerance	10% increase in grain yield over improved check (Thialack II) with >40 ppm Fe
Early maturity, high yielding **hybrids** for adaptation to better endowed environment	Target: central and northern part of the peanut basin agro-ecological zone	Early 70–90 days / (350–800 mm/annum)	**Biotic stresses:** Downy mildew resistance, head miner **Abiotic stresses:** Drought, low soil fertility	Grain yield of > 2.0 tons/ha Downy mildew resistance Head length > 50 cm	High grain iron and zinc content. Striga tolerance Drought tolerance	15% increase in grain yield over improved check (Souna 3) with >40 ppm Fe

Chapter 2- Genomic approaches to build de novo elite breeding gene pools from locally-adapted landraces

2.1 Summary

Many nascent breeding programs aim to achieve genetic gain by crossing locally elite germplasm, but a lack of systematic approaches to develop elite gene pools from locally-adapted varieties hinders their progress. Motivated by the observation of undesirable transgressive segregation in presumed elite crosses in Senegalese cereal breeding programs, we designed approaches for de novo development of elite gene pools from locally-adapted landrace-derived germplasm. We first define two types of “elite" germplasm: *cis*-elite, phenotypically similar and genetically homogeneous for locally-adapted traits ("acquired traits"); versus *trans*-elite, phenotypically similar, but genetically heterogeneous for acquired traits. Next, we defined two genomic approaches for de novo inference of elite gene pools: population-based genotypic inference (PGI) and QTL-based genotypic inference (QGI), and compared to a family-based phenotypic inference (FPI) approach. Using simulations that trace the evolution from locally-adapted landraces to elite breeding lines, we evaluate the effectiveness of these strategies in nascent forward breeding programs. QGI accurately and cost-effectively identifies both *cis*- and *trans*-elite pairs, regardless of the underlying trait architecture, while PGI is less sensitive when trait architecture is oligogenic. Over ten cycles of phenotypic recurrent selection, programs based on *cis*-elite crosses consistently outperformed those based on *trans*-elite crosses for genetic gain. The findings highlight the value of trait genetic architecture knowledge for elite gene pool development and provide a practical roadmap for elite germplasm development in modernizing breeding programs.

2.2 Introduction

Intercrossing elite lines to develop improved cultivars is at the core of modern plant breeding (Rutkoski, 2019). Despite widespread use of "elite" in plant breeding literature, however, there remains no universally accepted definition of *elite*, nor systematic guidance on how to develop *elite* germplasm (Acquaah, 2012; Bernardo, 2020; Falconer, 1996; Falk, 2010). The term is variously applied to denote superior phenotypic performance, the potential to produce superior progeny, and/or genetic similarity within breeding populations (Table S1). Advanced breeding programs, such as public-sector barley programs or private-sector corn breeding programs in North America and Europe, have typically managed narrow elite gene pools (Bornowski et al., 2021; Rasmusson & Phillips, 1997). By contrast, emerging breeding programs often begin with much more diverse local andraces that are well-appreciated by local farmers and consumers, but the lack of clearly defined elite gene pools makes it difficult to design elite-by-elite crosses (Kane et al., 2022). Conversely, attempts to import "elite" germplasm from established breeding programs in different geographies has often failed because that germplasm is maladapted to local environments and cultural needs of smallholder farmers (Walker & Alwang, 2015). Therefore, new approaches are needed to develop elite germplasm de novo from locally-adapted landrace germplasm.

Genetic gain requires deliberate population improvement strategies that balance stabilizing selection to maintain favorable traits and directional selection to drive progress toward breeding goals (Falconer, 1996; Hill, 2010). Stabilizing selection helps preserve adaptive genetic variation by maintaining key allele frequencies, while directional selection promotes the accumulation of favorable alleles to improve target traits (Walsh & Lynch, 2018a). The interplay between these selective forces underlies the genetic architecture of elite germplasm and influences long term response to selection (Hill, 2010; Walsh & Lynch, 2018a). Breeding programs that manage this balance effectively sustain additive genetic variance, which according to the breeder’s equation, is critical for maximizing genetic gain (Lush, 1943).

Landrace origins are shaped by a combination of farmer selection preferences, local environmental pressures, and seed exchange networks, which can promote genetic heterogeneity through diverse adaptation or lead to homogeneity under isolation or intensive selection (Nelimor et al., 2020; Rattunde et al., 2021). In crops such as sorghum, millet, and cowpea, this diversity is a critical resource, but it complicates the definition and development of elite germplasm. Phenotypic similarity among lines may mask underlying genetic heterogeneity

(Brown, 1989; Crossa et al., 2017), and individual-based selection risks overlooking this cryptic variation (R & B, 1981). In contrast, established programs have often used family-based selection to capture hidden structure and stabilize elite populations (Bernardo, 2020; Comstock & Robinson, 1948). These contrasting strategies highlight the need for emerging programs to adopt more deliberate frameworks that integrate diversity and structure in the development of elite gene pools.

Population genetic theory on parallel evolution is one promising area to inform frameworks for elite gene pool development. When distinct lineages from related founder populations are exposed to similar environmental pressures, they often evolve toward similar phenotypes, a common outcome in evolution known as parallel evolution (Wood et al., 2005). The evolutionary processes leading to genomic parallelism when multiple lineages independently evolve toward a similar phenotype in similar environments (Bohutínská et al., 2021). Individuals from different closely related founding populations adapt in similar ways under the same conditions with different genetics (Blount et al., 2018; Whiting et al., 2022). In addition, the pressure and direction of selection and genetic drift applied to individual quantitative trait loci (QTL) can give rise to alleles with an antagonistic effect for the trait, and thus increase the extent and frequency of transgressive segregation from crosses in domesticated populations (Dickinson et al., 2003). While parallel evolution has been studied in domestication and local adaptation, its implications for crop breeding remain largely underexplored.

Here, we present a framework for de novo elite gene pool development in emerging breeding programs through the use of molecular markers. Our objective was to identify accurate, efficient, and cost-effective strategies for inferring elite genetic relationships within diverse germplasm pools to optimize crossing decisions and accelerate genetic gain. To this end, we compared a phenotypic inference approach, family-based phenotypic inference (FPI), with two new genomic inference approaches, population-based genotypic inference (PGI) and QTL-based genotypic inference (QGI), evaluating their effectiveness across traits with differing genetic architectures. Our findings demonstrate that genotypic inference approaches can be successfully applied to establish elite gene pools, with particular strengths depending on the underlying genetic basis of target traits. These results highlight an opportunity to deploy genetic knowledge for de novo development of elite gene pools, offering practical guidance for emerging breeding programs aiming to balance eliteness with genetic diversity.

2.3 Results

2.3.1 Unexpected transgressive segregation from presumed elite breeding crosses

To assess the status of eliteness within the germplasm of emerging breeding programs, we analyzed progenies from recent elite-by-elite crosses made by breeders, along with effectiveness of ongoing trait introgression efforts. The motivating example for this study was the observation of transgressive segregation in putative elite-by-elite Senegalese sorghum and pearl millet breeding programs (Figures 1A–D). In pearl millet, a cross between the high-yielding, early-flowering variety Souna 3 and Thialack 2, valued for its long panicles, resulted in only 33% of F_5 progenies flowering within the acceptable range (Figures 1A), undermining expectations for combining favorable traits (Figures 1C). A similar outcome occurred with a sorghum cross between two late-flowering lines, Nguinthe and Grinkan, intended to enhance panicle size (Figures 1B and 1D). In both cases, the broad segregation for flowering time reduced the number of candidates for advancement, and therefore the effective population size (Figures 1C–D).

2.3.2 Conceptual model for genetic heterogeneity in elite germplasm: cis- versus trans-eliteness

In many emerging breeding programs particularly across sub-Saharan Africa early efforts over the past century focused on collecting and stabilizing locally adapted landraces into pure line varieties (Figure 2A). These presumed elite lines were developed to consolidate key adaptive traits, such as flowering time, while enhancing agronomic performance traits like yield and disease resistance (Figure 2A). Traits for which an acceptable phenotypic value has already been captured during this process are referred to here as “acquired traits” (under stabilizing selection). By contrast, “desired traits” (under directional selection) are those for which breeding efforts aim to change the phenotypic value, often to meet farmer preferences or evolving environmental demands. Crosses are often made based on phenotypic similarity without accounting for underlying genetic relatedness or genotypic complementarity. We hypothesize that excessive transgressive segregation observed in crosses of similar parent lines may be explained by cryptic genetic heterogeneity shaped by parallel evolution in isolated breeding efforts (Figure 2B–C).

Thus, we propose that two contrasting types of elite relationships may exist: *cis*- versus *trans*-elite. In the *cis*-elite cross scenario, repeated selection and genetic drift within locally adapted germplasm may result in elite lines that share similar genotypes at acquired traits (Figure 2B). Crosses between such genetically homogeneous lines typically produce narrow phenotypic

distributions, with most progeny mostly resembling the parents in both trait phenotype and genetic makeup. This facilitates efficient selection and helps maintain a robust effective population size (Figure 2B). By contrast, the *trans*-elite cross scenario would occur after a history of parallel adaptation of populations with the same phenotype, but different genotypes at acquired traits. Here, elite lines derive from different ancestral lineages and carry distinct haplotypes, with alleles in repulsion phase (Figure 2C). Crosses between such divergent lines can lead to extensive transgressive segregation due to the recombination of incompatible or novel allelic combinations (Figure 2C). While this genetic novelty may occasionally yield superior individuals for desired traits, it more often results in progeny with unacceptable values for acquired traits. This would lower the proportion of acceptable progeny, reducing effective population size, and therefore, genetic gain.

2.3.3 Proposed approaches to infer cis- versus trans-elite relationships

We propose three approaches to infer elite line relationships, each grounded in a distinct genetic rationale and suited to varying resource capacities:

- *Family-based Phenotypic Inference*: FPI would involve crossing elite lines in a diallel or partial diallel design and phenotypically evaluating the resulting segregating progeny (e.g., F_2 to F_5 generations) (Figure 3A). The theoretical expectation is that *cis*-elite pairs, which share similar haplotypes, will produce narrow phenotypic distributions, while *trans* -elite pairs will yield broader segregation due to greater allelic divergence. Since this approach requires only classical breeding techniques (crossing, phenotyping, and selection) it may have been how (intentionally or unintentionally) classical breeders in the twentieth century developed elite gene pools. This approach provides inferences without genomic or molecular tools but is time-consuming and resource-intensive.

- *Population-based Genotypic Inference:* PGI would leverage genome-wide marker data to infer population structure among elite lines (Figure 3B). We applied Discriminant Analysis of Principal Components (DAPC), a multivariate method that clusters individuals based on genetic variation without assuming Hardy-Weinberg equilibrium. DAPC maximizes between group variation while minimizing within group variance, allowing the identification of genetically homogeneous clusters. The rationale is that elite lines belonging to the same genetic cluster (i.e., same subpopulation) are more likely to share ancestral haplotypes, especially in genomic regions under selection. Such lines are

thus inferred as having a *cis*-elite relationship, reducing the likelihood of unexpected segregation when crossed.

- *QTL-based Genotypic Inference:* QGI would use genetic similarity within known QTL that underlie acquired traits that are essential for acceptability (e.g. flowering time, plant height, or grain color in cereals) (Figure 3C). Identity-by-state (IBS) matrices are used to compare elite lines based on shared allelic states at locus-specific SNPs within trait-associated QTLs. This approach provides a more precise assessment of genetic compatibility by directly examining variation at underlying loci rather than genome-wide background. Elite lines that share high IBS values at relevant QTLs are considered *cis*-elite pairs with respect to the studied trait, whereas those with divergent alleles enabled allelic recombination increase the risk for transgressive segregation and therefore inferred *trans*-elite pairs. QGI is theoretically the most targeted and informative approach, but it depends on prior knowledge of QTL or quantitative trait nucleotides (QTNs) and access to high-quality genotypic data.

To evaluate the potential effectiveness of these approaches we simulated the steps in the evolution of nascent breeding programs (Figure 2) and applied each of the approaches (Figure 3) to the simulated breeding programs (Figure 4; see Materials and Methods for details). These steps included: (1) creating diverse, locally-adapted landraces and defining the trait (Figure 4A), (2) purifying landraces to develop initial local varieties (Figure 4B), (3) cross-breeding initial varieties to establish a forward breeding program (Figure 4C), (4) inter-crossing progeny of *cis*- vs. *trans*-elite relationships to initiate a recurrent selection program (Figure 4D), and (5) evaluating outcomes for the nascent breeding program (Figure 4E).

2.3.4 Family-based phenotypic inference is simple but limited to simple acquired trait architecture

We evaluated the effectiveness of FPI in distinguishing *cis*- and *trans*-elite relationships by analyzing F_2 progeny distributions and validating predictions against F_5 outcomes under simulated genetic architectures involving 2, 4, or 20 QTLs (Figures 5A–C). However, across all QTL scenarios, FPI showed no consistent phenotypic pattern reliably distinguishing the two elite types (Figure 5A for QTL = 2, and Figure 5B for QTL = 4, and Figure 5C for QTL = 20). To apply the approach, we introduced a breeder-defined threshold: crosses were classified as *trans*-elite if more than 75% of F_2 individuals fell outside the predefined acceptable range of

phenotypes. Under this criterion, FPI correctly identified elite type in 70% of cases when 4 QTLs were involved, suggesting that the approach can be moderately effective when clear trait thresholds are established. Despite its simplicity and intuitive appeal, performance of FPI declined under more complex genetic architectures, highlighting its limitations as trait architecture complexity increases. Thus, while FPI offers a straightforward framework for elite-type inference, its accuracy is contingent on both trait complexity and specific phenotypic thresholds.

2.3.5 Population-based genotypic inference can be effective for polygenic acquired trait architecture

Next we determined the accuracy of using the PGI approach. PGI failed to accurately detect *trans*-elite relationships at a lower number of loci (Figures 5D–E). For QTL numbers of two and four, F_5 progeny from *trans*-elite crosses showed a narrow distribution for some runs (Figure 5D for QTL = 2; Figure 5E for QTL = 4). Many families of inferred *trans*-elite crosses have narrow distributions leading to erroneous conclusions and thus unpredictable distribution in subsequent generations. In contrast, putative *cis*-elite crosses gave the expected distribution across all 100 runs (Figure 5D for QTL= 2; Figure 5E for QTL = 4). The accuracy of inference for *trans*-elite relationships improved as the number of QTL underlying the acquired trait increased (Figure 5F). At twenty QTL for the acquired trait, *trans*-elite combinations were detected as accurately as *cis*-elite pairs and the accuracy of PGI was similar to QGI (Figure 5F).

2.3.6 QTL-based genotypic infers elite type across a range of trait genetic architectures

To determine whether QGI in known QTLs can be used to accurately infer *cis*- versus *trans*-elite pairs, we analyzed the phenotypic distribution in F_5 generations for three genetic architecture scenarios. The progeny of *cis*-and *trans*-elite pairs inferred using QGI follows this prediction (Figures 5G–I). Across all three genetic architecture scenarios, *cis*-elite crosses produce offspring with a narrow distribution for the acquired trait (flowering time in days from sowing to flowering), almost all in the acceptable range of phenotypes. In contrast, *trans*-elite crosses generated offspring with a broad phenotypic distribution for flowering time when this trait is considered to have oligogenic architecture (Figure 5G for QTL = 2, Figure 5H for QTL = 4, and Figure 5I for QTL = 20). In fact, when fewer high-effect variants segregate, the many possible allelic combinations give rise to a wider range of phenotypes for the trait. Thus, most of the progeny from these *trans*-elite crosses are outside the acceptable range of phenotypes and the number of loci is fixed to twenty (Figure 5I). When the acquired trait has multiple small effect

variants (nQTL = 20), the phenotypic distribution is still broader compared to the progeny of the *cis*-elite crosses. However, 94% of the offspring remain within the acceptable range of phenotypes as small-effect allele deviates the recombinants less from the initial average (Figure 5K).

2.3.7 QTL-based genotypic inference provides highest accuracy for elite-type inference

To assess the accuracy of elite-type inference, we compared phenotypic variance within progeny from *cis*- and *trans*-elite crosses across the three approaches (FPI, PGI, and QGI) under simulated genetic architectures involving 2, 4, and 20 loci (Figure 5J). A reliable inference approach is expected to yield low phenotypic variance among *cis*-elite progeny and high variance among *trans*-elite progeny. FPI produced the highest overall phenotypic variance across all scenarios, with limited distinction between *cis*- and *trans*-elite crosses (Figure 5J). While *trans*-elite crosses showed higher variance than *cis*-elite crosses at 2 and 4 loci, the variance among *cis*-elite progeny was unexpectedly high. At 20 loci, variance in *cis*-elite crosses remained elevated, suggesting that FPI becomes increasingly unreliable as trait complexity increases. PGI performed better than FPI, revealing significant differences in variance between *cis*- and *trans*-elite crosses (*p*-values: 0.0015, 0.0131, and 0.0007 for 2, 4, and 20 loci, respectively) (Figure 5J). However, the magnitude of variance separation was modest compared to QGI. QGI consistently outperformed the other approaches, showing highly significant differences in phenotypic variance between *cis*- and *trans*-elite progeny at all locus levels ($p < 0.001$) (Figure 5J). It also demonstrated a clear trend: variance decreased with increasing locus number for both cross types, but remained distinctly lower for *cis*-elite and higher for *trans*-elite crosses.

2.3.8 QTL-based genotypic inference is the most cost-effective approach when trait loci are known

We compared the cost-effectiveness of two genotypic inference approaches PGI and QGI against the FPI approaches, assuming prior knowledge of trait-associated loci (Table S3). When the number of loci was low (≤4), QGI was the most cost-effective, with estimated costs of $75 and $85 per line for 2 and 4 loci, respectively, compared to $125 for PGI. However, QGI costs rose with increasing locus number and surpassed PGI when 20 loci were involved ($165 per line). In contrast, when trait loci were unknown and required discovery via genome wide association studies (GWAS) or linkage mapping, QGI became significantly more expensive $843 and $670 per line, respectively exceeding both PGI and FPI (Table S4). Under comparable

assumptions, FPI, which relies on crossing and phenotyping, incurred an estimated cost of $904 per line per environment (Table S3). Overall, PGI emerged as the most cost-effective approach when no prior knowledge of trait loci was available.

2.3.9 Genetic gain under recurrent selection is more rapid for cis-elite crosses

To evaluate the impact of *cis*- versus *trans*-elite crossing strategies on genetic gain for a desired trait, we simulated 10 cycles of phenotypic recurrent selection, evaluating the performance of F_5 progenies. We initially hypothesized that *cis*-elite crosses, despite potentially compromising genetic diversity for the acquired trait, would still match or exceed genetic gain in the desired trait (e.g., yield potential). However, the relative performance of elite gene pool crossing strategy depended strongly on the subpopulation origin of the parental lines.

For the acquired trait, progenies from *cis*-elite crosses maintained a stable population mean and narrow phenotypic variance across all cycles, indicating strong trait stabilization (Figure 6A). *Trans*-elite progeny exhibited broader variation early on but gradually stabilized (Figure 6B), while total genetic variance declined over time for both strategies, with *trans*-elite crosses showing an initial peak due to recombination (Figure 6C). For the desired trait, outcomes differed based on whether *cis*-elite parents originated from the same or different subpopulations. When *cis*-elite parents shared the same subpopulation, *trans*-elite crosses yielded greater genetic gain over time (Figure 6D–F). Although *trans*-elite progenies started with broader phenotypic variance (Figure 6E), this variance declined over cycles and contributed to a stronger selection response. Genetic variance for the desired trait was higher in *trans*-elite crosses during the early cycles, supporting this advantage (Figure 6F). However, when *cis*-elite parents originated from different subpopulations, they outperformed *trans*-elite crosses in final genetic gain (Figure 6G–I). These crosses combined subpopulation diversity while retaining elite backgrounds, achieving both trait stability (Figure 6G) and higher final population means (Figure 6H), with comparable variance reduction over time (Figure 6I).

2.4 Discussion

2.4.1 A modernized theoretical and applied framework for eliteness in breeding programs

Many nascent breeding programs, particularly in developing countries, are implementing individual-based phenotypic selection (IPS; Table 1, Figure 3C) in crosses of locally-adapted landrace-derived varieties, but seeing little success in terms of genetic gain or varietal adoption (Walker & Alwang, 2015). Breeding efforts in low-income resource contexts have historically

relied on phenotypic selection without access to genomic tools, so lines with locally-adaptive acquired traits are often designated as *elite* based on phenotypic performance alone. Our findings show that crosses between such lines can produce unexpected segregation patterns, indicative of underlying genetic heterogeneity (Figure 1, 2, 4). These outcomes suggest hidden genetic heterogeneity and contradict the notion that elite status equates to genetic homogeneity, a premise often overlooked or taken for granted in breeding literature (Falk, 2010) (Table S1). To address this gap, we develop a framework for defining and operationalizing *eliteness* in nascent breeding populations. By explicitly separating these two categories of traits (acquired versus desired), breeding strategies can be better aligned to harness recombination potential for desired traits while maintaining necessary levels of stability for acquired traits (Figure 2, 6). While we developed the theory and approaches with a focus on nascent programs in developing countries, similar genetic heterogeneity may exist within established networks of public programs, such as for winter wheat in the US Great Plains (Ayalew et al., 2020; Tessele et al., 2025), or even mature commercial hybrid maize programs (Technow et al., 2021).

Family selection remains a fundamental strategy in plant breeding, particularly in more mature programs where elite lines are extensively tested through their progeny performance (Acquaah, 2015). The theoretical basis of this approach assume that phenotypic differences among families can be attributed to both genetic divergence and the compatibility of parental lines (Walsh & Lynch, 2018). The evaluation of FPI sought to clarify the extent to which these assumptions hold under varying genetic architectures, particularly in distinguishing between *cis*- and *trans*-elite relationships (Figure 5). The rationale behind FPI aligns with classical family selection theory: that phenotypic distributions in progeny reflect underlying genetic structure, with narrower distributions expected from genetically similar (*cis*-elite) crosses, and broader segregation arising from genetically distinct (*trans*-elite) combinations (Falconer, 1996; Walsh & Lynch, 2018b). However, our results reveal a disconnect between this theoretical expectation and empirical outcomes (Figure 5). Across simulated trait architectures ranging from 2 to 20 QTLs, FPI failed to consistently differentiate elite types based on phenotypic variance alone. This finding underscores a key limitation of conventional family-based inference namely, its sensitivity to genetic complexity and its dependence on phenotypic thresholds that may not fully capture the genetic basis of trait expression.

Our findings build on previous theoretical work emphasizing the interplay between within- and between-family genetic variance in shaping selection responses (Fehr, n.d.;

Simmonds, 1996; Walsh & Lynch, 2018), and call attention to the need for more robust approaches that integrate genomic information with phenotypic evaluation to identify compatible genotypes. Incorporating a breeder-defined cutoff (i.e., ≥25% of F_2 individuals outside an acceptable range of phenotypes) improved FPI's predictive utility to 70% accuracy in oligogenic scenario, suggesting a moderate practical value when trait phenotype is well characterized and thresholds are carefully chosen (Figure 5). However, as the number of QTLs increased, predictive accuracy declined, highlighting the challenges of applying simple inference models to complex, polygenic traits. Interestingly, while we modeled additive acquired traits, the fact that they are under stabilizing selection (i.e. there is a concave trait-fitness function) means that the acquired trait loci are epistatic in terms of fitness (i.e. the fitness effect of an allelic substitution depends on the alleles at other loci). Therefore, one interpretation of our findings is that *cis*-eliteness represents genotypes on the same epistatic peak of an adaptive landscape while *trans*-eliteness represents genotypes on a different epistatic peak (Wade, 1992; Wright, 1932).

2.4.2 Use genomics findings to guide de novo breeding elite gene pools development

The relevance of the QGI approach in this study is closely tied to the nature of the acquired trait, which is controlled by a small number of major-effect loci. This aligns with longstanding debate in crop improvement regarding the value of trait mapping. While some have argued that mapping efforts offer limited utility for improving highly quantitative traits (Bernardo, 2008; Rutkoski, 2019), and that genome-wide selection may yield faster gains (Bernardo & Yu, 2007), our findings suggest that when the architecture of an acquired trait is oligogenic or monogenic, trait mapping is both relevant and impactful. Under such conditions, QGI targeting known loci offered high accuracy, biological relevance, and cost-effectiveness for elite type inference (Figures 5D–F, Table 1). These findings reinforce the value of causal gene identification in enhancing breeding efficiency, particularly in stress-resilience contexts where traits like flowering time or photoperiod sensitivity are key (Guitton et al., 2018).

Programs wishing to deploy QGI could leverage extensive trait mapping in orphan crops such as sorghum, pearl millet, and cowpea, which are targeted by many emerging breeding programs across sub-Saharan Africa and South Asia. In these crops, acquired traits like flowering time, plant height, and drought tolerance have been successfully mapped due to rich phenotypic variation and segregation within traditional landraces (Boukar et al., 2016; Faye et al., 2021; Haussmann et al., 2002). Cloned loci such as *Ma1* (SbPRR37) and *Dw1* in sorghum (Murphy et al., 2011; Yamaguchi et al., 2016), and QTLs identified for flowering time and plant

height in pearl millet (Varshney et al., 2017) and cowpea (Lo et al., 2018), could be used in QGI to define *cis*-elite gene pools. The QGI approach could leverage this trait knowledge to robustly distinguish *cis*- and *trans*-elite relationships, offering speed and cost advantages that are essential for adoption by resource-limited breeding programs.

Our simulations of PGI provide qualified support for earlier proposals to use population structure to identify breeding gene pools from locally-adapted landraces and global breeding germplasm (Faye et al., 2021). While this strategy is sound when polygenic trait alleles are already fixed (Figure 2B), our results show that PGI performs poorly for oligogenic traits (Figures 5D–E), where trait loci are few and may not align with principal genomic patterns. PGI's reliance on overall genomic relationships makes it vulnerable when causal variation is confined to specific regions. Additionally, the accuracy of PGI was influenced by the clustering strategy DAPC performance varied based on the number of PC axes selected, a well-known issue that can lead to false subpopulation detection (Cullingham et al., 2023). Although we used the conservative *optim.a.score* method, this can overestimate dimensions under moderate-to-high gene flow. In scenarios where only a few QTLs underlie the trait, such false structure could obscure true elite-type relationships. Nevertheless, as QTLs represent a small fraction of genome -wide data used for clustering, the limitations of PGI under oligogenic control are consistent with theoretical expectations.

By contrast, QGI demonstrated superior performance by targeting genomic regions linked to specific trait loci. It was robust across trait architectures and retained accuracy even when PGI failed (Figures 5G–I). Though PGI and QGI converged in performance for polygenic traits, QGI's specificity made it a more reliable inference tool overall. Notably, QGI maintained high inference accuracy for monogenic and oligogenic acquired traits, provided that trait loci had been mapped (Figures 5D–F). However, as the number of loci increases, inference precision can decline (Figure 5J), due to the increased number of plausible haplotypes and variation in IBS scores across genome regions (Henden et al., 2018; Waples et al., 2019). IBS-based inference assumes SNPs are in or near causal loci; thus, QGI's effectiveness depends on accurate trait mapping and marker resolution (Clark et al., 2013). In summary, our findings suggest that molecular marker tools can effectively guide elite gene pool development in emerging breeding programs, especially when the genetic architecture of key traits is well understood. QGI, in particular, stands out as a cost-effective and biologically grounded strategy for inferring *cis*- and *trans*-elite relationships especially when targeting a small number of high-priority traits. As

breeding programs continue to face urgent demands with limited resources, approaches like QGI offer a practical path toward accelerating crop improvement.

Finally, our simulations demonstrate that the genetic diversity and the performance of *cis* -elite crosses depend strongly on the subpopulation origin of the parental lines. When *cis*-elite parents are drawn from the same subpopulation, they exhibit reduced genetic variability, which limits transgressive segregation for genetic yield potential and the likelihood of generating superior phenotypes for desired traits (Figure 6D–F). This finding is consistent with prior findings that crossing closely related lines restricts segregational variance, leading to reduced long-term genetic gain particularly when the trait architecture is oligogenic and polygenic (Bernardo, 2020; Clark et al., 2013). In contrast, *cis*-elite crosses between parents from different subpopulations retain the elite background while introducing meaningful genetic divergence, thereby increasing recombination opportunities and enhancing selection response (Figure 6G–I). This suggests that such *cis*-elite crosses combine stability for acquired traits with the genetic divergence typically targeted for desired traits. *Trans*-elite progenies, while initially more variable (Figure 6E), also benefited from increased genetic variance early in selection, which enabled greater short-term gains in cases where *cis*-elite parents were genetically uniform. These findings emphasize that population structure within elite pools is a critical design consideration for maximizing genetic gain under recurrent selection (Falconer, 1996; Swarup et al., 2021).

2.5 Conclusion

This study highlights that presumed elite lines in emerging breeding programs often harbor hidden genetic heterogeneity, leading to unexpected transgressive segregation when crossed. By distinguishing acquired from desired traits, we demonstrate that elite-by-elite crosses vary in outcome depending on the genetic relationship of parents. While costs will vary across programs and over time, in our initial estimates QTL-based genotypic inference (QGI) emerged as the most accurate and cost-effective method for identifying *cis*- and *trans*-elite relationships, especially when trait loci are known. Our simulations show that *cis*-elite crosses between parents from different subpopulations strike the best balance maintaining stability in acquired traits while promoting recombination for desired traits thus optimizing long-term genetic gain. These findings provide a practical extension to the breeder's equation, showing how informed elite cross design can enhance selection efficiency without increasing population size.

2.6 Materials and methods

2.6.1 Simulating diverse locally-adapted landraces

The MaCS software (Chen et al., 2009), implemented via the AlphaSimR package (Faux et al., 2016; Gaynor et al., 2021; Gorjanc et al., 2016), was used to generate an ancestral founding population of 500 outbred lines, setting the species history to GENERIC. This parameter enables the simulation of haplotype sequences shaped by historical changes in population size (Faux et al., 2016; Gaynor et al., 2021). Specifically, the effective population size was modeled to change over time: starting at 10,000 individuals 100,000 generations ago and decreasing stepwise to 6,000, 1,500, 500, and finally 100 individuals in the present, thus mimicking periods of reduced genetic diversity. The genome was modeled with 7 chromosomes, consistent with pearl millet (2n = 2x = 14 chromosomes), and each chromosome had 1,000 random segregation sites. The physical and genetic lengths were set to 1×10^{8} base pairs and 1 Morgan, respectively. Haplotype sequences were generated using the Markovian Coalescent Simulator (MaCS), with a mutation rate of 2.5×10^{-8} per base pair (Chen et al., 2009). The founding population was simulated independently 100 times using the same parameters to mimic 100 independent breeding programs (Figures 3 and 4).

For each simulation, two categories of traits were defined in the founding population: a moderately heritable acquired trait (e.g., flowering time or plant height), and a lowly heritable desired trait (e.g., genetic yield potential). These traits were modeled based on phenotypic values observed in the Senegalese pearl millet local germplasm. The acquired trait represented by flowering time was simulated with a mean of 55 days, variance of 10 days, and narrow-sense heritability (h^2) of 0.6. The acceptable phenotypic range for flowering time was defined by setting minimum and maximum values at 52.5 and 57.5 days, respectively, to guide stabilizing selection. The desired trait grain yield was modeled with a mean of 900 kg/ha and a low heritability of 0.1, reflecting its complexity and high environmental influence in landraces. Grain yield potential for each individual was modeled based on its phenotypic value for the acquired trait, with penalties applied to those whose trait values fell outside the acceptable range of phenotypes. The further an individual's phenotype deviates from the optimal mean and acceptable range for flowering time, the greater the penalty applied to its grain yield potential.

Three oligogenic genetic architecture scenarios were simulated for the acquired trait, assigning 2, 4, or 20 quantitative trait loci (QTLs). In contrast, the desired trait was modeled as polygenic, controlled by 100 QTLs. All QTLs were independently and randomly distributed

across the chromosomes. For each QTL scenario, equal additive effect sizes were simulated using an external function, defined by the following equations (see Figures 3 and 4).

$$addeff = rep(value, total\_nQTL) \quad \text{Equation 1}$$

$$value = ((max.t1 - min.t1) + 3) / total\_nQTL \quad \text{Equation 2}$$

Where total_nQTL is the number of loci controlling the trait, max.t1 and min.t1 represent the maximum and minimum values of the acceptable range of phenotypes. These equations ensure that the additive effects are scaled to maintain trait variability across different QTL numbers.Five subpopulations of 100 individuals each were derived from an ancestral founder population (Figure 3). Within each subpopulation, individuals underwent random mating for 50 generations. At each generation, the top 10% of individuals based on grain yield and falling within an acceptable range for flowering time were selected and recycled for the next round of mating. This selection and mating scheme was applied independently within each subpopulation, mimicking local adaptation to similar environmental conditions while allowing genetic drift to shape allele frequencies over time.

2.6.2 Simulating purification of landraces to create initial local varieties

Following 50 generations of selection and random mating within each of the five subpopulations, the resulting lines were merged to form a single assembly germplasm representing a broad base of locally adapted diversity. From this assembly germplasm, 200 individuals were selected to represent distinct landrace lines, capturing both phenotypic stability in the acquired trait and variation in the desired trait. These selected lines formed the founding breeding population for subsequent improvement. To generate initial local varieties suitable for selection and further evaluation, each of the 200 lines was advanced through selfing using the single seed descent (SSD) method, producing homozygous elite lines. This purification step served to stabilize trait expression and create a panel of fixed lines representing locally adapted landraces with elite potential.

2.6.3 Simulating cross-breeding of initial varieties without eliteness inference

To simulate conventional improvement strategies without prior inference of elite relationships (individual-based phenotypic selection; IPS), the top 10% of highest-yielding lines among 200 purified lines were designated as elite. Trait donor lines were selected from within this elite group, each carrying beneficial alleles for specific agronomic traits of interest (e.g., drought tolerance, disease resistance). A forward breeding approach was employed: selected local varieties were crossed with these trait donors, and the resulting progeny underwent

phenotypic screening to identify individuals demonstrating superior performance for both introduced and target traits. Progeny exhibiting favorable phenotypes were advanced through standard breeding pipelines, with top-performing individuals selected as candidates for multi-location and advanced field trials. This strategy reflects typical breeding practices in the absence of prior elite-type inference and served as a comparative baseline for evaluating the effectiveness of structured gene pool approaches.

2.6.4 Phenotypic inference of elite gene pools

FPI was deployed using a half-diallel mating design to evaluate all possible combinations among selected elite parental lines. The F_2 progenies derived from each cross were assessed to characterize the distribution of phenotypic variation within families. These distributions were used to infer the relationships (*cis* vs. *trans*) and define gene pools. Crosses that produced F_2 populations with narrow phenotypic distributions were inferred to involve parents in a *cis* relationships, while those with broader distributions were interpreted as *trans* relationships, reflecting greater segregation variance. To validate these initial inferences, we examined the distribution of phenotypic variation within F_5 progenies derived from the same crosses.

2.6.5 Genotypic inference of elite gene pools

For QGI, we extracted the genotype information in the QTL and then calculated the identity By State (IBS) for each pair of individuals for each marker position using the cluster function daisy() from the R package (Figure S3). Pairs of individuals with the highest IBS indices equal to 1 were considered *cis* elite pairs (Figure S3). In contrast, pairs of individuals with the IBS indices inferior to 1 were considered *trans* elite pairs (Figure S3).

For PGI, a data matrix of molecular markers across all segregating sites was used to first identify clusters using successive K-means with the function find.clusters(). A first discriminant analysis of principal components (DAPC) was performed with these clusters using the dapc() function and then used optim.a.score() to define the optimal number of Principal Components (PCs). A second DAPC was performed using the defined optimal number of PCs. Euclidean distance was used to infer *cis*- and *trans*-elite relationships based on clustering (Figure S3). The Euclidean distance between individuals between and within clusters using the following equation:

$$\text{Euclidean}(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \quad \text{Equation 3}$$

The pairs of individuals with the lowest and highest Euclidean distance within the cluster with the highest proportion of successful reassignment were inferred as having *cis*-elite and *trans*-elite

relationships, respectively (Figure S3).

2.6.6 Analysis of the phenotypic distribution of offspring

Five pairs of putative *cis*- and *trans*- elite combinations were crossed respectively to sequentially generate one F_1 progeny and selfed to obtain 100 F_2 progeny then, 100 F_5 individuals per single-seed descent from each pair (Figures 3 and 4). The phenotypic distribution for flowering time of F_5 progeny of *cis*- and *trans*- elite crosses was analyzed using a histogram in *ggplot2* R package, respectively. The accuracy of elite type prediction was therefore assessed based on the phenotypic distribution. While a narrow distribution is expected for F_5 progeny of *cis*- elite crosses, a narrow distribution is expected for F_5 progeny of *trans* elite crosses (Figures 1C-D).

2.6.7 Comparing costs of de novo gene pool strategies

We estimated the costs of defining breeding gene pools for 50 elite lines using QGI and PGI and compared them to FPI. Cost estimates were based on the millet breeding program budget in Senegal during the 2016/2017 season (Table S2). Figure S1 outlines the sequence of steps considered for each strategy. The phenotypic approach involved crossing the 50 elite lines with four required control lines (depending on the trait) using a half-diallel mating design. F_1 seeds were harvested from each cross and then advanced to F_5 generation progenies, following the procedures previously described (Figure 4). The phenotypic distribution of the F_5 progenies was then analyzed for the trait of interest.

For the genotypic approaches, we considered two scenarios: (i) major QTLs already known, and (ii) QTLs unknown and requiring discovery. The latter scenario includes additional steps such as forward genetic methods GWAS and biparental QTL mapping as well as reverse genetic approaches like allele mining and comparative genomics supported by functional genomics (Figure S1). In this study, cost estimates were based solely on forward genetics. For GWAS, we estimated the cost of genotyping and phenotyping a diversity panel of 300 lines across two sites over two years. For biparental QTL mapping, we simulated the development and evaluation of 250 F_2 individuals, including genome-wide genotyping and trait phenotyping. Once major QTLs were identified, the next steps included genotyping the 50 elite lines at these specific regions for allele screening and computing similarity matrices for QGI analysis. To estimate PGI costs, we assumed genome-wide genotyping of the 50 elite lines followed by PGI analysis for elite clustering.

Genotyping costs were estimated at $20 per line for genome-wide data and $5 per line for each QTL-specific assay based on previous genotyping activities conducted in Senegalese the sorghum breeding program. Phenotyping costs were estimated at $64 per line, based on the same 2016/2017 Senegalese pearl millet program budget. The experimental design used to derive phenotypic cost estimates was an alpha-lattice design (7 × 13), with three replications. Each elementary plot consisted of two rows of 8 mounds, flanked by two border rows, with 90 cm spacing between rows and between mounds. Costs included all field operations in a rainfed environment from land preparation to harvest covering labor, inputs, and equipment (Table S2).

2.6.8 Simulating intercrossing to initiate recurrent selection program

To initiate the recurrent selection program, three additional acquired traits were simulated, resulting in a total of four acquired traits and one desired trait modeled in the ancestral founding population. The purification process to create initial local varieties was conducted as previously described. QGI was used exclusively to infer *cis*- and *trans*-elite relationships among parental lines. These inferred elite parent pairs were then used to initiate a phenotypic recurrent selection (PRS) program. In cycle 0 (C0), two scenarios were simulated independently: one starting with five *cis*-elite pairs and the other with five *trans*-elite pairs. From each cross, one F_1 plant was generated and selfed to produce 100 F_2 individuals via single seed descent, resulting in a total of 500 F_5 individuals per scenario (Figure 4). These F_5 individuals were advanced through subsequent generations via the same method. At each cycle, selection was based on performance for the desired trait (e.g., grain yield), while maintaining acquired traits within an acceptable range of phenotypes. This applied directional selection on the desired trait and stabilizing selection on the acquired traits. The phenotypic traits were modeled as independent traits in the simulation. A total of ten selection cycles were simulated for each scenario to evaluate the long-term effectiveness of the phenotypic recurrent selection strategy. The outcomes of the nascent breeding program were evaluated by tracking phenotypic trends for both acquired and desired traits across selection cycles using *ggplot2* in R. In addition, total genetic variance for these traits was extracted and visualized to assess the impact of selection on genetic diversity over time.

Acknowledgements

This study is made possible through funding by the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through grants from American People provided to the United States Agency for International Development (USAID) under cooperative

agreement number AID-OAA-A-13-00047 and 7200AA-19LE-00005. The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the US Government. Support was also provided by the Gates Foundation through the grant “Green Evolution - Accelerating Dryland Cereals Improvement for Africa (INV-053669)” to GPM.

Chapter 2 References

Acquaah, G. (2012). Principles of Plant Genetics and Breeding (2 edition). Wiley-Blackwell.

Ayalew, H., Sorrells, M. E., Carver, B. F., Baenziger, P. S., & Ma, X.-F. (2020). Selection signatures across seven decades of hard winter wheat breeding in the Great Plains of the United States. *The Plant Genome*, *13*(3), e20032. https://doi.org/10.1002/tpg2.20032

Bernardo, R. (2008). Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Science*, *48*(5), 1649–1664. https://doi.org/10.2135/cropsci2008.03.0131

Bernardo, R. (2020). *Breeding for quantitative traits in plants* (Third edition). Stemma Press.

Bernardo, R., & Yu, J. (2007). Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science*, *47*(3), 1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science*, *362*(6415), eaam5979. https://doi.org/10.1126/science.aam5979

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., Slotte, T., & Kolář, F. (2021). Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences*, *118*(21), e2022713118. https://doi.org/10.1073/pnas.2022713118

Boukar, O., Fatokun, C. A., Huynh, B.-L., Roberts, P. A., & Close, T. J. (2016). Genomic Tools in Cowpea Breeding Programs: Status and Perspectives. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00757

Brown, A. H. D. (1989). Core collections: A practical approach to genetic resources management. *Genome*, *31*(2), 818–824. https://doi.org/10.1139/g89-144

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, *19*(1), 136–142. https://doi.org/10.1101/gr.083634.108

Clark, S. A., Kinghorn, B., & Van Der Werf, J. H. (2013). *Comparisons of Identical by State and Identical by Descent Relationship Matrices Derived from SNP Markers in Genomic Evaluation*. Association for the Advancement of Animal Breeding and Genetics (AAABG). https://rune.une.edu.au/web/handle/1959.11/14477

Comstock, R. E., & Robinson, H. F. (1948). The Components of Genetic Variance in Populations of Biparental Progenies and Their Use in Estimating the Average Degree of Dominance. *Biometrics*, *4*(4), 254–266. https://doi.org/10.2307/3001412

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., Campos, G. de los, Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Cullingham, C., Peery, R. M., & Miller, J. M. (2023). A roadmap to robust discriminant analysis of principal components. *Molecular Ecology Resources*, *23*(3), 519–522. https://doi.org/10.1111/1755-0998.13724

Dickinson, H. G., Hiscock, S. J., Crane, P. R., Rieseberg, L. H., Widmer, A., Arntz, A. M., & Burke, J. M. (2003). The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1434), 1141–1147. https://doi.org/10.1098/rstb.2003.1283

Falconer, D. S. (1996). *Introduction to Quantitative Genetics*. Pearson Education.

Falk, D. E. (2010). Generating and maintaining diversity at the elite level in crop breeding. *Genome*, *53*(11), 982–991. https://doi.org/10.1139/G10-081

Faux, A.-M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., Hearne, S. J., Gonen, S., & Hickey, J. M. (2016). AlphaSim: Software for Breeding Program Simulation. *The Plant Genome*, *9*(3). https://doi.org/10.3835/plantgenome2016.02.0013

Faye, J., Maina, F., Akata, E., Sine, B., Diatta, C., Mamadou, A., Marla, S., Bouchet, S., Teme, N., Rami, J., Fonceka, D., Cisse, N., & Morris, G. (2021). A genomics resource for genetics, physiology, and breeding of West African sorghum. *The Plant Genome*, *14*. https://doi.org/10.1002/tpg2.20075

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

Gorjanc, G., Jenko, J., Hearne, S. J., & Hickey, J. M. (2016). Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genomics*, *17*, 30. https://doi.org/10.1186/s12864-015-2345-z

Guitton, B., Théra, K., Tékété, M. L., Pot, D., Kouressy, M., Témé, N., Rami, J.-F., & Vaksmann, M. (2018). Integrating genetic analysis and crop modeling: A major QTL can finely adjust photoperiod-sensitive sorghum flowering. *Field Crops Research*, *221*, 7–18. https://doi.org/10.1016/j.fcr.2018.02.007

Haussmann, B. I. G., Mahalakshmi, V., Reddy, B. V. S., Seetharama, N., Hash, C. T., & Geiger, H. H. (2002). QTL mapping of stay-green in two sorghum recombinant inbred populations. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *106*(1), 133–142. https://doi.org/10.1007/s00122-002-1012-3

Henden, L., Lee, S., Mueller, I., Barry, A., & Bahlo, M. (2018). Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLoS Genetics*, *14*(5), e1007279. https://doi.org/10.1371/journal.pgen.1007279

Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 73–85. https://doi.org/10.1098/rstb.2009.0203

Lo, S., Muñoz-Amatriaín, M., Boukar, O., Herniter, I., Cisse, N., Guo, Y.-N., Roberts, P. A., Xu, S., Fatokun, C., & Close, T. J. (2018). Identification of QTL controlling domestication-related traits in cowpea (Vigna unguiculata L. Walp). *Scientific Reports*, *8*(1), 6261. https://doi.org/10.1038/s41598-018-24349-4

Lush, J. L. (1943). Animal breeding plans. *Animal Breeding Plans.*, *Edn 2*. https://www.cabdirect.org/cabdirect/abstract/19440100562

Murphy, R. L., Klein, R. R., Morishige, D. T., Brady, J. A., Rooney, W. L., Miller, F. R., Dugas, D. V., Klein, P. E., & Mullet, J. E. (2011). Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences*, *108*(39), 16469–16474. https://doi.org/10.1073/pnas.1106212108

Nelimor, C., Badu-Apraku, B., Garcia-Oliveira, A. L., Tetteh, A., Paterne, A., N'guetta, A. S.-P., & Gedil, M. (2020). Genomic Analysis of Selected Maize Landraces from Sahel and Coastal West Africa Reveals Their Variability and Potential for Genetic Enhancement. *Genes*, *11*(9), 1054. https://doi.org/10.3390/genes11091054

R, H. A., & B, M. F. J. (1981). *Quantitative genetics in maize breeding. 1st ed.* https://agris.fao.org/search/en/providers/123819/records/64735c8a2c1d629bc97c8113

Rattunde, F., Weltzien, E., Sidibé, M., Diallo, A., Diallo, B., Brocke, K. vom, Nebié, B., Touré, A., Traoré, Y., Sidibé, A., Diallo, C., Diakité, S., Bretaudeau, A., & Christinck, A. (2021). Transforming a traditional commons-based seed system through collaborative networks of

farmer seed-cooperatives and public breeding programs: The case of sorghum in Mali. *Agriculture and Human Values*, *38*(2), 561–578. https://ideas.repec.org//a/spr/agrhuv/v38y2021i2d10.1007_s10460-020-10170-1.html

Rutkoski, J. E. (2019). A practical guide to genetic gain. In *Advances in Agronomy* (Vol. 157, pp. 217–249). Elsevier. https://doi.org/10.1016/bs.agron.2019.05.001

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science*, *61*(2), 839–852. https://doi.org/10.1002/csc2.20377

Technow, F., Podlich, D., & Cooper, M. (2021). Back to the future: Implications of genetic complexity for the structure of hybrid breeding programs. *G3 Genes|Genomes|Genetics*, *11*(7). https://doi.org/10.1093/g3journal/jkab153

Tessele, A., Morris, G. P., Akhunov, E., Johnson, B. E., Clinesmith, M., & Fritz, A. K. (2025). *Interchromosomal Linkage Disequilibrium Analysis Reveals Strong Indications of Sign Epistasis in Wheat Breeding Families* (p. 2025.05.16.654524). bioRxiv. https://doi.org/10.1101/2025.05.16.654524

Varshney, R. K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A., Srivastava, R. K., Chitikineni, A., Fan, G., Bajaj, P., Punnuri, S., Gupta, S. K., Wang, H., Jiang, Y., Couderc, M., … Xu, X. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature Biotechnology*, *35*(10), 969–976. https://doi.org/10.1038/nbt.3943

Wade, M. J. (1992). Sewall Wright: Gene interaction and the shifting balance theory. *Oxford Surveys in Evolutionary Biology*, *8*, 35–35.

Walker, T. S., & Alwang, J. (2015). *Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa*. CABI. https://iaes.cgiar.org/spia/publications/crop-improvement-adoption-and-impact-improved-varieties-food-crops-sub-saharan

Walsh, B., & Lynch, M. (2018a). *Evolution and Selection of Quantitative Traits*. Oxford University Press.

Walsh, B., & Lynch, M. (2018b). Family-Based Selection. In B. Walsh & M. Lynch (Eds.), *Evolution and Selection of Quantitative Traits* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198830870.003.0021

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, *28*(1), 35–48. https://doi.org/10.1111/mec.14954

Whiting, J. R., Paris, J. R., van der Zee, M. J., & Fraser, B. A. (2022). AF-vapeR: A multivariate genome scan for detecting parallel evolution using allele frequency change vectors. *Methods in Ecology and Evolution*, *13*(10), 2167–2180. https://doi.org/10.1111/2041-210X.13952

Wood, T. E., Burke, J. M., & Rieseberg, L. H. (2005). Parallel genotypic adaptation: When evolution repeats itself. *Genetica*, *123*(1–2), 157–170. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2442917/

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. 6th International Congress of Genetics*, *1*, 356–366.

Yamaguchi, M., Fujimoto, H., Hirano, K., Araki-Nakamura, S., Ohmae-Shinohara, K., Fujii, A., Tsunashima, M., Song, X. J., Ito, Y., Nagae, R., Wu, J., Mizuno, H., Yonemaru, J.-I., Matsumoto, T., Kitano, H., Matsuoka, M., Kasuga, S., & Sazuka, T. (2016). Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. *Scientific Reports*, *6*, 28366. https://doi.org/10.1038/srep28366

Chapter 2 Figures and Tables

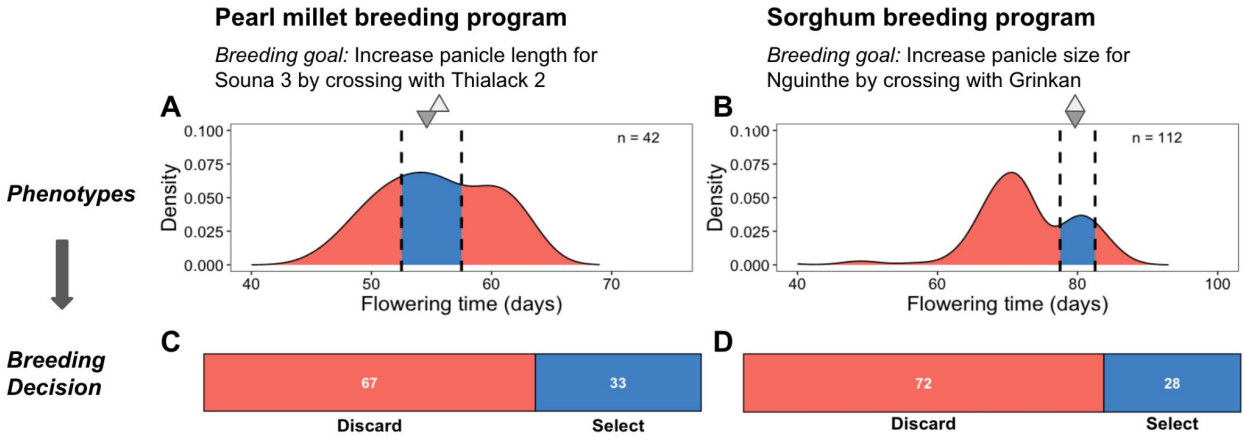


Figure 1. Undesirable transgressive segregation for an acquired trait in progeny from presumed elite-by-elite crosses substantially reduces the number of candidates for selection. Phenotypic distribution of flowering time in progeny from presumed elite-by-elite crosses in millet (A; F_5 progeny) and sorghum (B; F_7 progeny) breeding programs in Senegal. For pearl millet, Souna 3 was crossed with Thialack 2 to improve panicle length while maintaining earliness and high yield. For sorghum, Nguinthe was crossed with Grinkan to improve panicle size while maintaining lateness. Flowering time is an acquired trait in these programs, since the parents (triangles) are within the acceptable range of phenotypes (black dotted lines) according to the breeding product profiles. For both crops, substantial transgressive segregation was observed, with few progeny within the acceptable range (blue shading in A and B), and most progeny falling outside (red shading in A and B). Accordingly, the percentage progeny of discarded (red in C and D) versus selected (blue in C and D) progeny was high, with only 33% and 28% of the progeny for pearl millet and sorghum, respectively, selected as acceptable with respect to flowering time.

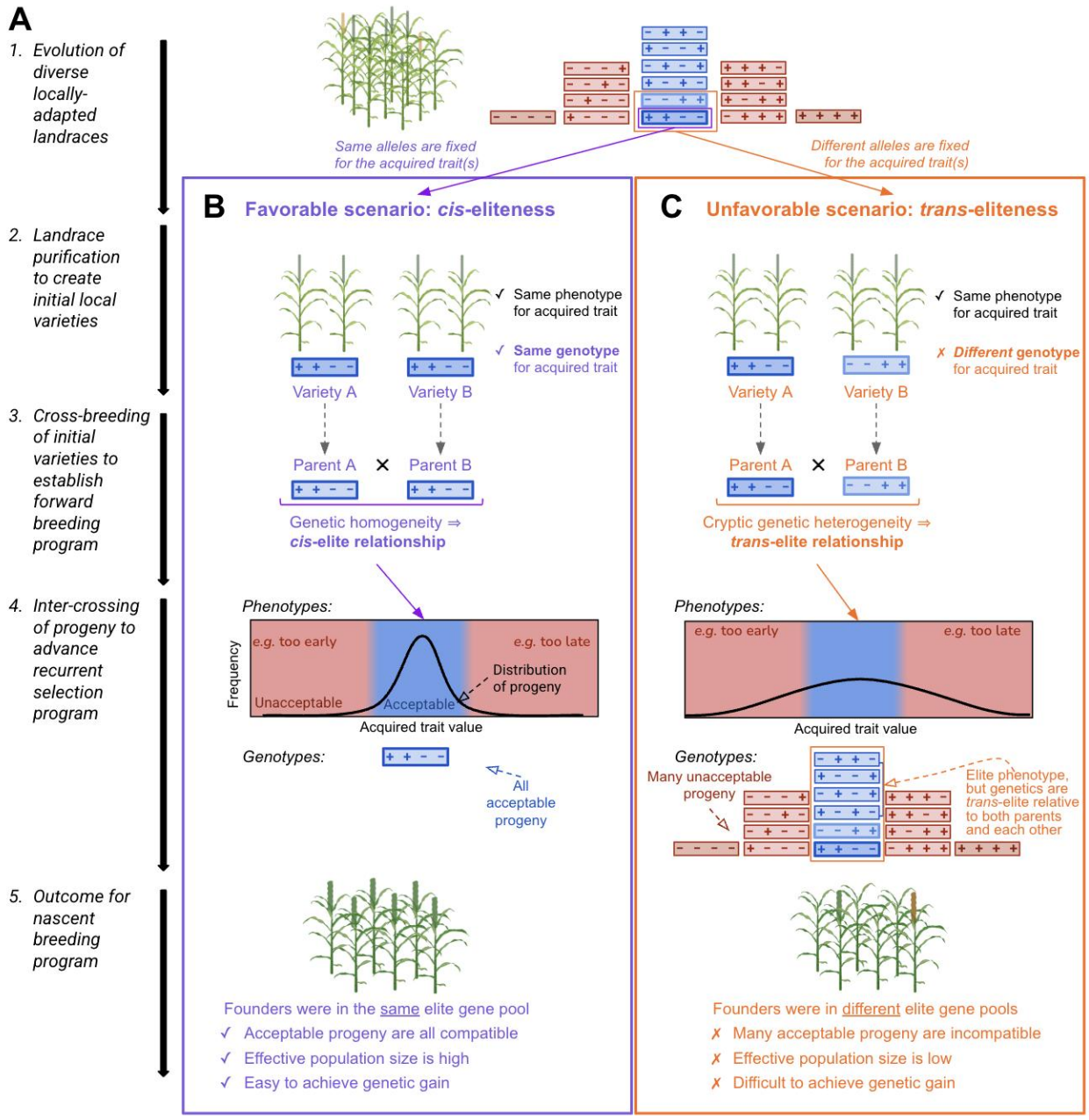
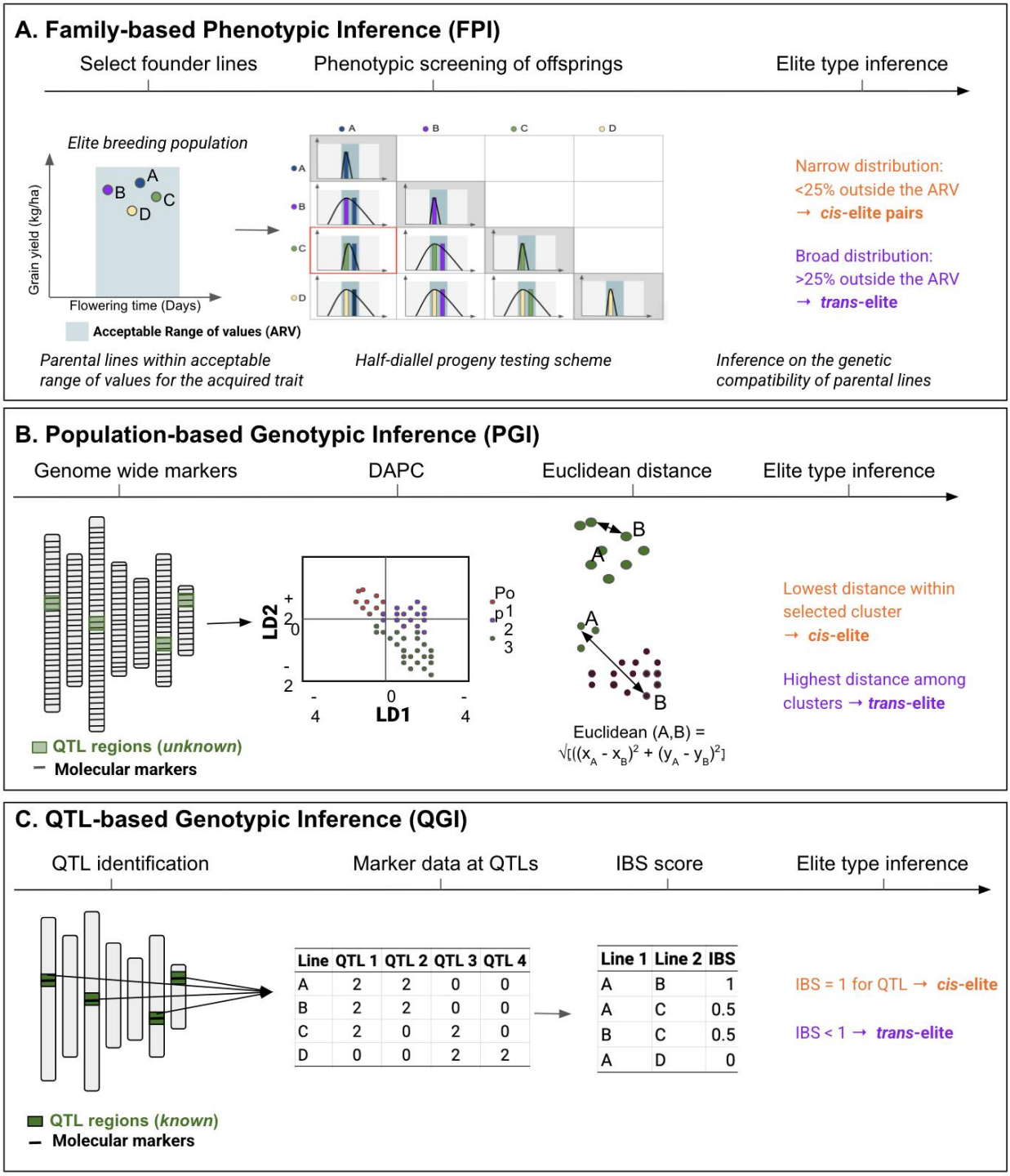


Figure 2. Evolutionary scenarios that lead to *cis*- versus *trans*-elite relationships in nascent breeding programs. (A) Schematic overview of the breeding trajectory, starting from landrace collection and parallel adaptation (Step 1), followed by landrace purification (Step 2), initial crossing of locally-elite lines (Step 3), intercrossing in recurrent breeding programs (Step 4), and culminating in the outcomes observed in nascent breeding programs (Step 5). Two distinct evolutionary scenarios are considered at each step: (B) *cis*-eliteness, characterized by genetic homogeneity among elite parents, and (C) *trans*-eliteness, characterized by cryptic genetic heterogeneity. Each rectangle represents the diploid genotype of a line, with "+" and "–" indicating favorable and unfavorable alleles, respectively, across four representative QTL for an acquired trait. (Heterozygosity is ignored for simplicity). In the *cis*-elite scenario (B), genetically

homogeneous elite parents produce progeny with a narrow phenotypic distribution, where most offspring fall within an acceptable range. By contrast, the *trans*-elite scenario involves genetically heterogeneous elite parents, resulting in a broader phenotypic distribution, with many progeny falling outside the acceptable range and being discarded.



Line	QTL 1	QTL 2	QTL 3	QTL 4
A	2	2	0	0
B	2	2	0	0
C	2	0	2	0
D	0	0	2	2

Line 1	Line 2	IBS
A	B	1
A	C	0.5
B	C	0.5
A	D	0

Figure 3. Schematic overview of approaches to inference *cis*- and *trans*- eliteness. (A) *Family-based phenotypic inference:* Elite lines were crossed using a half-diallel mating design. The *cis*- or *trans*-elite relationships were inferred from the phenotypic distribution of F_2 progeny, and the accuracy of these inferences was evaluated using F_5 progeny. Narrow phenotypic distributions were indicative of *cis*-eliteness, while broader distributions suggested *trans*-eliteness. (B) *Population-based genotypic inference (PGI).* Whole-genome marker data were used to perform Discriminant Analysis of Principal Components (DAPC) and to calculate pairwise Euclidean genetic distances. Lineages from the same subpopulation, showing low genetic distances, were inferred to have *cis*-elite relationships, while those from different subpopulations, showing higher distances, were inferred to have *trans*-elite relationships. (C) *QTL-based genotypic inference (QGI).* Identity-by-state (IBS) scores were calculated using markers located within known QTL. A score of IBS = 1 was interpreted as evidence of *cis*-eliteness, while IBS < 1 was interpreted as *trans*-eliteness.

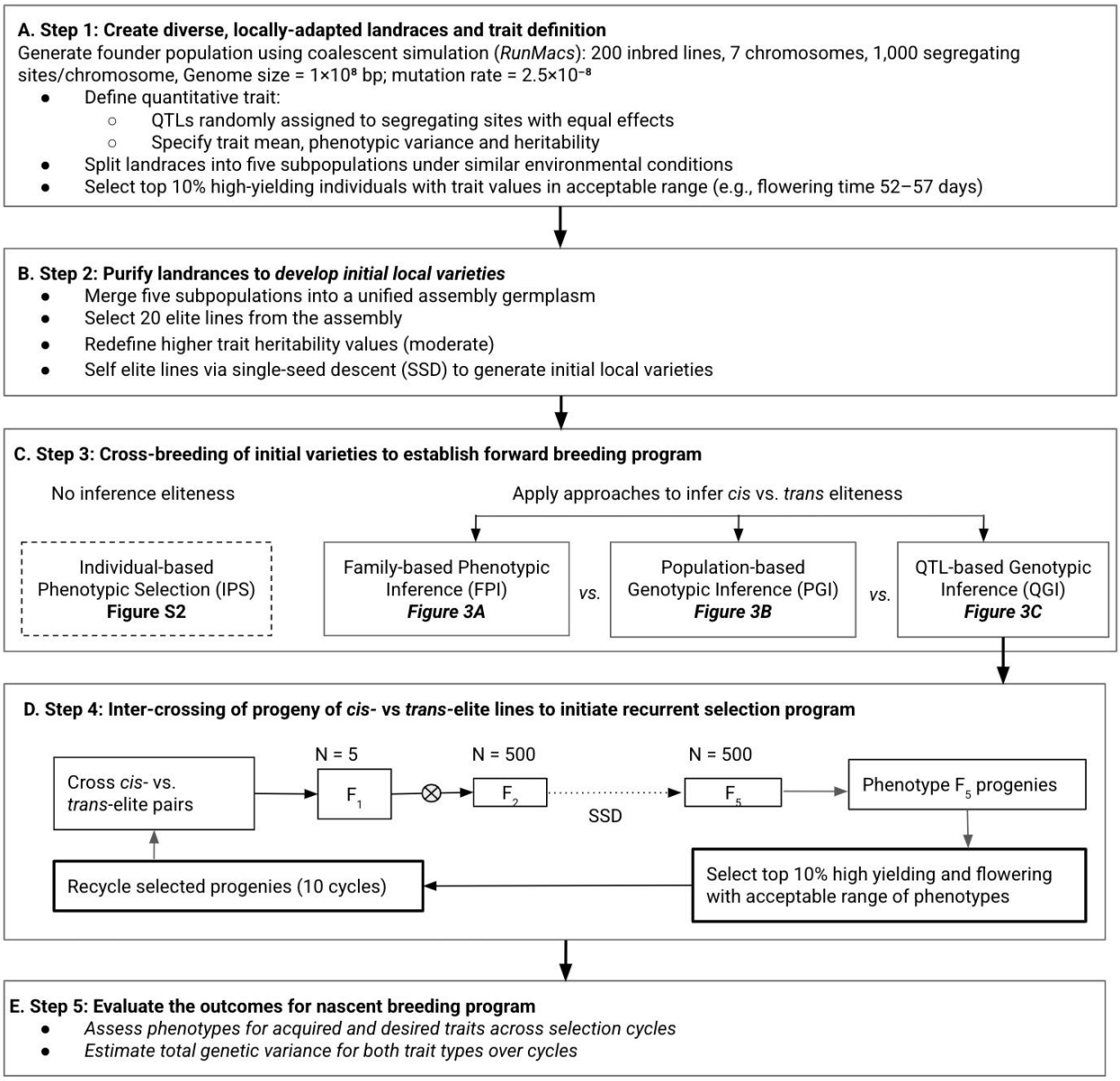


Figure 4. Simulation workflow for de novo creation, deployment, and evaluation of elite gene pools in nascent breeding programs. We conducted a five-step simulation process: (A) Generation of diverse, locally adapted landraces with defined trait architecture; (B) Purification and development of initial local varieties; (C) Cross-breeding to initiate a forward breeding program; (D) Intercrossing progeny with *cis*- or *trans*-elite relationships to launch phenotypic recurrent selection; (E) Evaluate the outcomes for nascent breeding program.

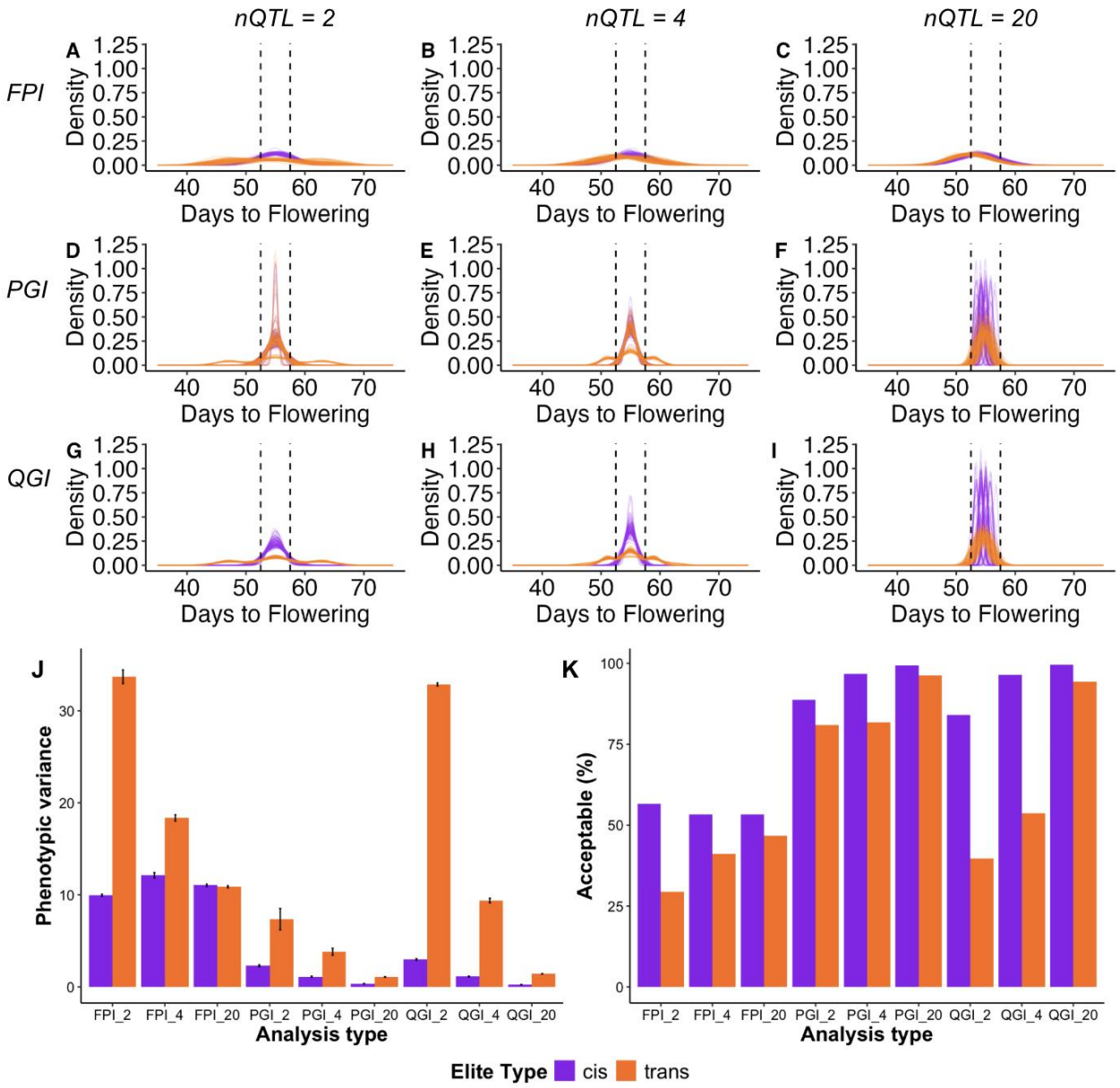

nQTL = 2
nQTL = 4
nQTL = 20
FPI
PGI
QGI
A
B
C
D
E
F
G
H
I
Density
Days to Flowering
J
Phenotypic variance
Analysis type
K
Acceptable (%)
FPI_2 FPI_4 FPI_20 PGI_2 PGI_4 PGI_20 QGI_2 QGI_4 QGI_20
Elite Type cis trans

Figure 5. QTL-based genotypic inference (QGI) is the most accurate approach for distinguishing *cis* and *trans*-elite relationships. Phenotypic distributions of F_5 progenies were compared under three alternative inference approaches: Family-based phenotypic inference (FPI) (panels A–C), QTL-based genotypic inference (QGI) (panels D–F), and Population-based genotypic inference (PGI) (panels G–I). Each panel represents results under different genetic architectures defined by the number of QTLs: 2, 4, or 20. Purple lines represent the distribution of progeny from *cis*-elite crosses (expected to be narrow), while orange lines represent trans-elite crosses (expected to be broader). Average phenotypic variance (J) and acceptable percentage (K) for the simulated acquired trait (flowering time) in F_5 progeny was computed for each approach (FPI, PGI, QGI) and QTL (2, 4, 20) scenario, labeled as: FPI_2, FPI_4, FPI_20, PGI_2, PGI_4, PGI_20, and QGI_2, QGI_4, QGI_20. For each scenario, 500 individuals were phenotyped, and the phenotypic variance was averaged across 100 simulation replicates.

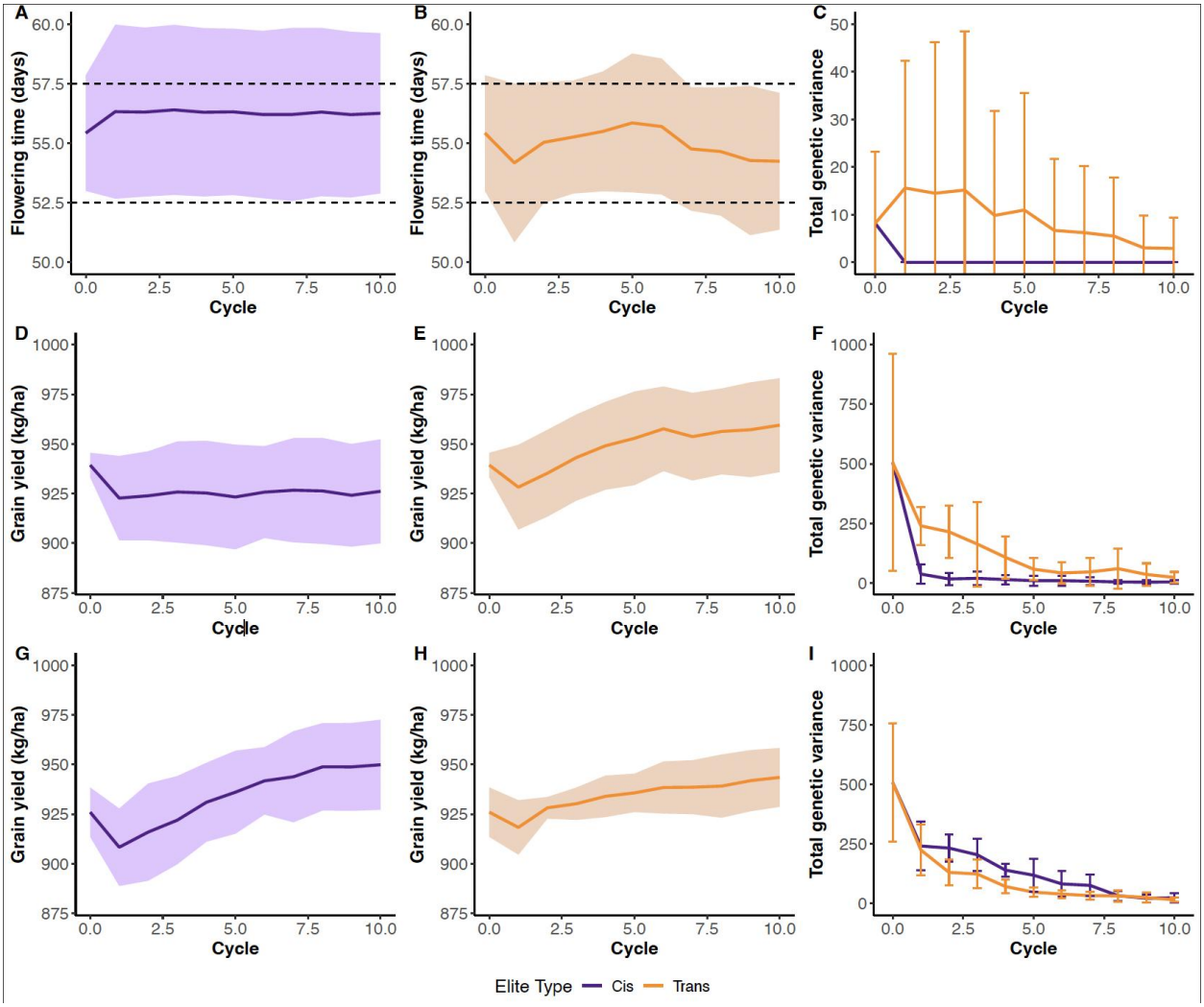


Figure 6. Genetic gain from *cis*-elite crosses surpasses *trans*-elite, but only when parents originate from different subpopulations. Average population mean and total genetic variance were tracked across selection cycles for progeny derived from *cis*- and *trans*-elite crosses. Panels A–C: Results for an oligogenic acquired trait, (A) population mean from *cis*-elite crosses, (B) population mean from *trans*-elite crosses, and (C) total genetic variance over cycles (*cis* vs. *trans*). Panels D–F: Results for a polygenic desired trait and *cis*-elite parents from the same subpopulation, (D) population mean from *cis*-elite crosses, (E) population mean from *trans*-elite crosses, (F) total genetic variance over cycles (*cis* vs. *trans*). Panels G–I: Results for a polygenic desired trait and *cis*-elite parents from the different subpopulations, (G) population mean from *cis*-elite crosses, (H) population mean from *trans*-elite crosses, and (I) total genetic variance over cycles (*cis* vs. *trans*).

Table 1. Comparative analysis of breeding approaches based on key evaluation criteria.

Approach	Infers eliteness	Uses genotypic information	Uses trait genetic information	Accurately infers *cis*-elite pairs	Accurately infers *trans*-elite pairs	Cost effectively infers elite type
IPS	✗	✗	✗	NA	NA	NA
FPI	✔	✗	✗	✗	✔	$$$
PGI	✔	✔	✗	✔	✗	$
QGI	✔	✔	✔	✔	✔	$

Checkmarks (✓) indicate positive performance, while crosses (✗) indicate absence or limitation in each respective category, relative cost is indicated using dollar signs ($ = low, $$$ = high). NA = not applicable.

Table S1. Conceptual ambiguity in the definition of *eliteness* in representative contemporary and classical plant breeding literature

Authors	Year	Definition [a]
Allier A, Teyssèdre S, Lehermeier C, Moreau L, Charcosset A,	2020	"Elite" not defined, but used 108 times in the text
Bernardo	2020	Eliteness not discussed
Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, Hagen T, Quinn M, Ng EH	2019	"**Elite germplasm** can be defined as a reproductively compatible set of genotypes disproportionately enriched for favorable alleles that improve breeding value (i.e., the ***mean performance of the progeny*** of a given parent) in a particular environment or market."
Cobb JN, Biswas PS, Platten JD	2018	"Elite" not defined, but used 61 times in the text
Lee	2015	"**Elite** – a crop line that has ***many genes for good agronomic traits*** that result in high yields in a particular environment."
Falk	2010	"Most breeding programs develop elite genotypes that are well adapted to the normal range of environmental conditions in the target production region. These **elite lines** have ***similar essential alleles*** for desirable end use characteristics, agronomics, disease resistance, and adaptation in the target region." "The new lines with fewer defects are the elite lines that form the core of most breeding programs, and recombination among the progeny of elite elite crosses is the source of most new cultivars released by modern breeding programs"
Poehlman, DA Sleper	1995	Not defined in glossary or index.
Fehr	1991	"The selection of parents for such characters generally involves selection of **elite germplasm** with the ***best performance for the character*** and the ***greatest genetic diversity available***" (pg. 122) "Elite" not indexed, but appears 15 times in the text

Allard	1959	Not defined in glossary or index. The "Choice of Parents" (pg. 116) refers to

[a] A summary of explicit or implicit definitions on eliteness, and/or references to eliteness in each reference. Emphasis (bold or italics) is ours, not the authors.

Table S2. Estimated field phenotyping budget per cycle for 404 pearl millet lines from the national core collection at the ISRA research station in Bambey, Senegal.

	Activities included	Cost (FCFA)	Cost (USD)
1. Field preparation	Plowing	25,000	$42
2. Inputs	Fertilizers, pesticides, and others products	150,000	$250
3. Labor	Temporary labor, Daily labor	3,200,000	$5,333
4. Material and supplies	Small material, Bags, Labels	200,000	$332
Total	-	3,525,000	$5,875
Total per line	-	**38,315**	**$63**
Total per plant	-	**8,725**	**$22**

Costs are presented in both the original currency (CFA Franc) and in US dollars, assuming a conversion rate of 1 USD = 600 FCFA.

Chapter 3 - *Phytochrome C*, a main locus among oligogenic variants underlying flowering ecotype divergence between early and late pearl millet

3.1 Summary

The flowering period is a key phenotypic trait targeted by pearl millet breeding programs for developing drought-tolerant varieties. A major challenge lies in tracking allelic combinations without a clear understanding of the genetic architecture underlying flowering time. This study aimed to uncover the genetic basis of flowering variation in West African pearl millet to inform the formation of elite breeding gene pools. First, we used classical genetic analysis to assess the homozygosity of major loci associated with flowering time in a national diversity panel. Next, we applied genome-wide association studies (GWAS) to dissect the genetic architecture of the trait. Population structure analysis was conducted to explore gene pool composition at local, national, and regional scales. Our results revealed an oligogenic architecture for flowering time across all diversity panels. Five SNPs were significantly associated with flowering time, including one colocalized with the *Phytochrome C (PhyC)* gene, a key regulator of photoperiod response. These SNPs represent strong candidates for marker development to support molecular breeding. We also observed that while ecotypes (early-flowering Souna and late-flowering Sanio) were clearly differentiated at the local scale, lineages from different ecotypes and countries were often grouped within the same genetic clusters at broader geographic levels. These findings provide valuable genomic resources and suggest that coordinated molecular breeding efforts, leveraging shared diversity across countries, could accelerate the development of drought-resilient pearl millet varieties while preserving traits valued by West African farmers.

3.2 Introduction

Flowering time is under strong pressure selection in both natural and breeding populations. The flowering period of individual plants determines the duration of the life cycle, therefore the frequency and intensity of abiotic and biotic stresses encountered by the population (Gaudinier & Blackman, 2020). On the one hand, developmental plasticity processes allow plants to sense new environmental stresses and adjust their flowering time in order to improve fitness and crop yield (Gaudinier & Blackman, 2020). Climate change indices were associated with phenotypic plasticity and a change in the allele frequency of high-effect variants in the flowering period related genes in local *Arabidopsis thaliana*, pearl millet, and maize populations among other plant species (Choquette et al., 2023; Saïdou et al., 2009; Scheepens et al., 2018). On the other hand, specific thresholds are assigned to flowering time in breeding product profiles to optimize grain yield based on targeted environments. The development of high-yielding pearl millet varieties from local populations with respect to the targeted flowering time required management of different gene pools within and between breeding populations, hence knowledge of the genetic architecture underlying the trait.

Escape from terminal drought episodes was ensured by shifting to earlier flowering pearl millet phenotypes in West Africa. Particularly in the Sahelian zone, severe terminal droughts have been recorded since the 1970s, with an imminent increase in their frequency according to all forecast models (Ragatoa et al., 2024). Faced with the reduced or altered distribution of precipitation, three different adaptation strategies have been highlighted in local varieties of millet, resulting from natural selection and the selection of farmers in West Africa: escape, avoidance, and tolerance. First, soil water absorption was optimized in landraces with rapid growth of primary roots and early colonization of deeper soil horizons, thus preserving high water potential (Passot et al., 2016). Second, a higher frequency of early-flowering phenotypes was observed in response to precipitation variability (Kane et al., 2022; Saïdou et al., 2009). Early flowering plants escape conditions unfavorable for seed maturation (Gaudinier & Blackman, 2020). Finally, the adoption of higher tillering and smaller grains allows local varieties to maintain their functionality in a water stress environment (Serba & Yadav, 2016).

Among these mechanisms, early flowering was likely one of the most consistent and determining factors for the adaptation of millet to drought-prone regions of the Sahel zone.

The early flowering phenotype may have been acquired in West African pearl millet landraces by high-effect causal variants (A. Faye et al., 2022; Saïdou et al., 2009). Despite the complex molecular architecture of the trait, divergence in the number of genetic components influencing the trait has been observed among grass species (Blümel et al., 2015). In sorghum, variation in phenology has been associated with variation in a few genes in a network conserved in Arabidopsis (Grant et al., 2023). In contrast, in maize and purple stiff brome, numerous small-effect QTLs colocalized with both conserved flowering time as well as peripheral genes were identified for flowering time in local populations (Buckler et al., 2009; Minadakis et al., 2024).In pearl millet, previous GWAS studies have shown evidence of oligogenic trait variation in Senegalese landraces and colocalization with *Phytochrome C* (Diack et al., 2020; A. Faye et al., 2022). Other studies have reported co-localization with major classes of flowering-related genes, including photoperiodic regulators, floral integrators, and florigens. The presence of these high-effect loci may enable faster trait adjustment over fewer selection cycles.

Knowledge of genetic architecture informs selection strategies and accelerates the development of varieties (Bernardo, 2008). Genetic architecture can be defined as the shape of the genetic contribution of genetics on the phenology of the trait which has different components among which the number of loci and the distribution of effect size are the most taken into account (Timpson et al., 2018). Depending on the number and effect size of the loci controlling the trait, different breeding strategies can be adopted to accelerate genetic gain (Muleta et al., 2018). Although trait architecture can be inferred a priori from available knowledge of other crop species, trait architecture depends on the evolutionary history of the genetic material. In a population with a recent selection history, size effect variations are not fixed (Orr, 2009). Therefore, the more potential fitness peaks that population has reached for a given trait, the more likely it is to obtain a polygenic trait (Orr, 2009). It is therefore essential to study the genetic architecture of the trait in germplasm closely related to the breeding population.

This study the genetic basis of flowering time variation in West African pearl millet. “Souna” and "Sanio” are the local designations of early and late flowering pearl millet ecotypes,

respectively in Senegal (Diack et al., 2020). We use the term *ecotype* to reflect both morphological and phenological divergence and its likely adaptive nature (Lowry, 2012). First, we test whether early flowering alleles have undergone homozygosity due to local adaptation. Second, we assess whether drought escape is mainly conferred by high-effect variants in known flowering genes. Using local, national, and regional diversity panels, we identify QTLs associated with flowering time and propose candidate markers to facilitate trait introgression into breeding programs.

3.3 Results

3.3.1 Phenotypic variation for flowering time among landraces at all sampling scales

To test whether multi-trait differentiation fits the standard model of Souna and Sanio ecotype divergence, we analyze the relationships between traits in the three panels. We used phenotypic data for lines from the local diversity panel (Figure 1A) , the national diversity panel (Figure 1B), and the regional diversity panel (Figure 1C) collected at different levels of sampling scales (Figures 1A–C). Phenotypic variations in flowering time for the two ecotypes were analyzed to determine whether there is true trait segregation between a priori classified early and late ecotypes. At the sampling scale level, significant statistical differences were observed between early and late ecotypes ($p < 0.001$ = ***), meaning that ecotypes are segregated for flowering time. However, a wider distribution between these two ecotypes was observed as the level of sampling scale decreased (Figures 1D–1F). In the local diversity panel, the Souna and Sanio ecotypes each have a narrow distribution, but no overlap was observed in the phenotypic values (Figure 1D). In contrast, in the regional diversity panel, there is no clear gap between the values of early and late ecotypes and an overlap was observed in the distribution (Figure 1F).

Next, we performed principal component analysis (PCA) and created scatterplots to visualize the consistency of relationships between traits at each level of the sampling scale (Figures 1G–1I). At all three sampling levels, two groups were observed for early and late genotypes (Figures 1G–1I). However, the distance between the two clusters decreases as the sampling scale increases. Both for the panel of the Niakhar collection and for the panel of the Senegalese collection, a clear distance between the early and late clusters is observed while in

the regional diversity panel, the early and late genotypes are distinguished but no distance between the two clusters was observed (Figures 1G–1I).

The relationship between phenotypic traits was characterized to test whether ecotype divergence was strong and consistent across these diversity panels. The relationship between the traits is consistent between the national and the regional diversity panels: the flowering period is not correlated either panicle length and plant height (Figures 1H–1I). In contrast, those two traits are consistently positively correlated in the three panels (Figures 1G–1I). In the local diversity panel, the flowering period is positively correlated with the height of the plant and the number of fertile tillers (Figure 1G). Panicle length is independent of flowering time and the number of fertile tillers, meaning that they do not influence each other in the same direction. This dynamic relationship between traits induces a weak and variable ecotypic divergence between the local sample level versus the national and regional level.

3.3.2 Abundant heritable variation at all scales for flowering time other agronomic traits

To determine whether phenotypic variation in flowering and key traits that drive adoption by farmers are related to genetic variation, we calculate the heritability of each trait in each diversity panel (Table 1). Regional diversity is distinguished by a lower heritability value (H^2= 0.82) for flowering compared to the local and national diversity panel, 0.99 and 0.98, respectively. The same trend was observed for plant height with a value of 0.73 in the regional diversity panel compared to 0.93 and 0.98 in the local and national diversity panels, respectively. Compared to other agronomic traits (flowering time, plant height), grain yield and tiller number show lower heritability values (0.67–0.87). These high heritability values (0.82–0.99) obtained for flowering time in three diversity panels lead to the conclusion that phenotypic variation in flowering time is strongly due to variation in the genetic components underlying the trait.

3.3.3 Flowering time also segregates within landraces

To determine whether the main flowering loci were fixed by either natural selection and/or farmer selection, we characterized the phenotypic distribution and variance for flowering time of the core parental lines in the national diversity and their own selves progenies. The two ecotypes were represented, Souna (SOUNA III, PLS38, PLS158, PLS224, PLS 100, PLS61,

PLS319, E21 and PLS312) and Sanio (E11 and PLS57), flowering before and after 60 days, respectively (Figure 2A). The shape of the distribution is highly variable for parental lines regardless of ecotypes, with some having a narrow distribution for E11, PLS38, PLS158, PLS61, and PLS312 in contrast to a broader distribution for PLS224, PLS100, PLS319, and E21 (Figure 2A). The selfed progenies of both ecotypes showed a wide and shifted distribution for flowering time (Figure 2A). For example, PLS61 flowered before 60 days, while 82% of its own progeny flowered late after 60 days. While the parent lines E11 and PLS57 all flowered after 60 days, 100%, and 63% of their own self progenies flowered before the 60 days threshold (Figure 2A). Additionally, offspring had higher variance compared to the parents (Figure 2B; $p < 0.001$ = ***). These observations suggest that flowering loci segregate in their own progenies because they are not fixed in parents.

3.3.4 Oligogenic variation for flowering time with colocalization with Phytochrome C

Genome-wide association studies were conducted for each panel to test the hypothesis of oligogenic variation in flowering time, with major variants colocalizing with known flowering time genes. In the local diversity panel, association peak is observed near the *PhyC* in the MLM (Figure 3A) but the *p*-value ($p = 10\text{-}^{5.8}$) above the FDR threshold but does not reach the Bonferroni threshold. MLMM detected a SNP (SOX387240.1_11493623) in chromosome 2 that colocalized with Phytochrome C orthologs found in the millet reference genome (Figure 3C). Lines homozygous for A alleles flower late around 83 days, while lines homozygous for T alleles flower much earlier, around 52 days (Figure 3D). The A/T heterozygous lines flowers 76 days closer than A/A suggest a dominant effect of the A allele.

In the national diversity panel, association peak is observed near the *PhyA*, *PhyC, and MADS14* in the MLM (Figure 4A), and the MLMM (Figure 4C) but the *p* value ($p = 10\text{-}^{4}$, 10-^{4}, and $10\text{-}^{3.2}$ respectively) above the FDR threshold but also does not reach the Bonferroni threshold. Two high-effect variants were detected by the MLMM: SNP SOX387240.1_45728909 in chromosome 2 and SNP SOX387244.1_203879098 in chromosome 6 (Figure 4C). These two SNPs do not colocalize with *a priori* candidate genes. SNP SOX387240.1_45728909 is within 100 kb of two genes encoding putative kinase-containing proteins (dpca2g109920 and

dpca2g109930). SNP SOX387244.1_203879098 colocalized in a ~1 kb window with the dpca6g439110 gene which codes for an uncharacterized protein. For SOX387240.1_45728909 only two allelic classes were observed. Individuals with A/C genotypes flowered late around 87 days, while A/A genotypes flowered early around 58 days (Figure 4E). For SNP SOX387244.1_203879098 the T/T genotypes flowered around 60 days, C/C genotypes flowered at 93 days and heterozygous C/T flowered around 82 days with a very large distribution (Figure 4F).

In the regional diversity panel, association peak is also observed the genomic regions encoded by *PhyA*, *PhyC*, and *MADS14* in the MLM (Figure 5A) and MLMM (Figure 5C) but the *p* value are also lower to the Bonferroni threshold.

3.3.5 Diversity clusters by ecotypes at the local scale but less at national or regional scale

To test whether gene pools are defined regardless of ecotypes and country of lineages, we analyzed the clustering of individuals in the three diversity panels separately (Figure 6). In the local diversity panel, two clusters were obtained separately, radically the Souna and Souna lineages (Figure 6A). In the national diversity panel, two groups were also obtained, but a mixture of ecotypes was found in each group (Figure 6B). Finally, the regional diversity panel shows a mixture not only at the level of ecotype but also of country of origin (Figure 6C). The six genotypic groups were defined with mixed early and late lines from different countries of origin. These results suggest that although gene pools at the local level are defined based on ecotype, the genetic components of flowering time do not define gene pools as sampling scale increases.

3.4 Discussion

3.4.1 Natural selection and farmer selection were not sufficient to fix the flowering loci in local landraces

Flowering time is one of the main adaptive traits targeted by natural selection and farmer breeding to improve yield under drought (Franks et al., 2007). The long history of pearl millet cultivation in West Africa since regional domestication events leads us to hypothesize that the main flowering locations would be fixed. In natural populations, allelic variants conferring novel

phenotypes with increased fitness in individuals are selected, and genetic drift increases allele frequencies as selection progresses (Orr, 2009). In Niger, an increased frequency of the early flowering allele in *PhyC* gene was observed between 1976 and 2003 in landraces that escaped more frequent terminal drought events (Vigouroux et al., 2011). Additional selection pressure, exerted by farmers, would lead to a progressive fixation of the selected alleles. Contrary to our expectations, provided strong evidence that the main flowering loci are still segregating within individual landraces in the national diversity panel (Figure 2). This suggests that the selection pressure from the both farmer and natural selection forces was not sufficient to fix the main loci underlying the flowering time phenotypic variation in the Senegalese landraces. This could be in part explained by gene flow from practices in the traditional farmer system with seed exchanges between farmers (Kane et al., 2022). In addition, the use of non-certified seeds (landraces and informally propagated modern varieties) selected solely on the basis of the phenotype of the parents which would promote cryptic genetic heterogeneity in subsequent seasons.

3.4.2 Accounting for population structure reduces false positives in GWAS across sampling scales

In GWAS studies, sampling in narrow scales reduced the confounding effect of population structure, which can lead to false-positive associations (Uffelmann et al., 2021). When individuals are genetically differentiated and the pooled populations have a different phenotypic distribution for the trait of interest, the assumption of the standard statistical method is biased (Sul et al., 2018). SNPs may have high *p*-values not because they are positively correlated with the trait, but because they are associated with population structure components (Uffelmann et al., 2021). This could explain the positive correlations with SNPs in unpredicted genes obtained as we increase the sampling scale (see Figure 3 vs. Figures 4–5). The assumption that individuals are unrelated could be corrected by incorporating information about population structure into the model and many alternative models have been proposed for this purpose (Sul et al., 2018).

We consider three models in this study that handle differently the potential confounding effects related to population structure. The MLM and MLMM model take population structure

into account. In the local diversity panel, one SNP in the candidate genes was associated with flowering time. However, in the national diversity panel, two unpredictable and positively correlated SNPs were found. These SNPs could be false positive but most likely in linkage disequilibrium with high effect variants (Sul et al., 2018; Uffelmann et al., 2021).

3.4.3 Allelic variation at the main Phytochrome C locus differentiates early and late ecotypes

GWAS of the local diversity panel provides strong evidence that Phytochrome C (PhyC) is a key photoperiodic regulator of flowering time in West African pearl millet. A high-effect SNP in the *PhyC* gene differentiates early- and late-flowering landraces from Niakhar (Figure 3), supporting the role of photoperiod perception in local adaptation and the ecotypic divergence between Souna (early) and Sanio (late) types (Figure 1). This is consistent with previous findings of *PhyC* associations with flowering time in Senegalese and Nigerien germplasm (Diack et al., 2020; A. Faye et al., 2022; Vigouroux et al., 2011). PhyC is a known essential light-sensitive protein which perceives far-red light signals and regulates photoperiodic flowering in grasses (Woods et al., 2014). In foxtail millet, the domestication-associated *SiPHYC* allele promotes flowering under short-day conditions by increasing sensitivity to daylength, while repressing flowering under long days through interaction with the photoperiodic pathway (Wang et al., 2022). Similarly, in sorghum, *SbPhyC* maps to the major flowering time locus *Ma5* and acts to inhibit flowering under long days in photoperiod-sensitive genotypes (Grant et al., 2023). In maize, *ZmPhyC* was shown to repress expression of *ZCN8*, the major florigen (Yang et al., 2013).

This conserved role of *PhyC* across grass species, together with the presence of a high-effect variant in pearl millet *PhyC*, supports the hypothesis that the Sanio ecotype retains an allele that represses flowering under long-day conditions. In contrast, the Souna ecotype likely carries a derived *PhyC* allele that reduces photoperiod sensitivity, enabling early flowering, a key adaptive trait for escaping terminal drought in the Sahel.

3.4.4 Next steps for the development of regional genomic resources

In this study, we showed that these two ecotypes were clearly differentiated based on flowering. However, ecotype divergence including other morphological traits such as plant

height and panicle length being variable and weak at regional level (Figure 1). From a breeding perspective, these results imply that selection on flowering time, panicle length, or plant height will induce different pressure patterns on any of the other traits (simultaneous increase for positive correlations; no effect for no correlation and antagonistic effect on trait values for negative correlations) depending on the population.

The lack of population structure based on either ecotype and countries of origin of lines in the regional diversity panel imply that shared gene pools exist in the regional germplasm (Figure 6). This finding will be of great interest to collaborative efforts to breed for drought-tolerant pearl varieties at the regional level. This encourages the development of regional breeding product profiles, as common donor lines could be identified for trait introgression within gene pools. However, these results are taken with caution because a fairly low SNP density (~10k SNPs) was used in the population structure analysis for the regional panel. This low quantity of high-quality SNPs is due to the lack of accessibility of information on the adapters used in the genotyping process (Kanfany et al., 2020). Acquiring genotypic higher quality genotypic data for this panel and re-performing the analysis would be essential to confirm these results.

SNPs with a high effect on flowering time variation were found in the different panels (Figures 3–5). These results provide additional evidence for an oligogenic genetic architecture of flowering time in West African germplasm. For such a genetic architecture, phenotypic or marker-assisted selection will accelerate elite development and/or maintenance of elite gene pools, therefore introgression of favorable alleles for other key adaptive traits (Chapter 2). Therefore, the next step would be to validate the causality of their candidate variants and develop a breeder-friendly marker that could be rapidly deployed in the breeding pipeline.

3.4 Conclusion

This study provides compelling evidence that flowering time variation in West African pearl millet is largely governed by an oligogenic architecture, with Phytochrome C (PhyC) emerging as a major locus differentiating early- and late-flowering ecotypes. The segregation of key flowering loci within landraces highlights ongoing genetic heterogeneity despite prolonged

natural and farmer-driven selection. Our results demonstrate that accounting for population structure is critical in GWAS to reduce false positives and accurately identify candidate loci. The presence of shared gene pools across regional germplasm panels suggests promising opportunities for coordinated breeding efforts leveraging common alleles for drought adaptation. These insights offer valuable genomic resources and a strong foundation for marker-assisted selection to accelerate the development and maintenance of elite pearl millet varieties. Future work should focus on functional validation of candidate variants and the development of breeder -friendly molecular markers to enhance precision breeding strategies.

3.5 Material and methods

3.5.1 Plant material

Three panels were used in this study: a local diversity panel, a national diversity panel, and a regional diversity panel subset of the Pearl Millet Inbred Germplasm Association (PMiGAP) panel. The study considered three levels: from a single location in Senegal (Niakhar, in the main rainfed agricultural area of Senegal), to a national coverage with different agroclimatic zones with the national diversity panel, and finally to a third level of sampling at regional scale including different countries across West Africa. The local and national diversity panels respectively comprised 91 and 110 landraces collected exclusively in Senegal, respectively. The regional panel included 138 open-pollinated varieties (OPV) and landraces from West African countries (Senegal, Burkina Faso, Mali, Niger, Nigeria, Ghana, Togo and Mauritania), all previously self-fertilized over 4 to 6 generations (Table S1) (Kanfany et al., 2020). Individuals were sampled from only one of the diversity panels and no overlap exists between these three diversity panels. The seeds of the regional diversity panel were shared by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT-Niger) and all collections were stored at the National Center for Agronomic Research (CNRA) in Bambey.

3.5.2 Field phenotyping

The lines from the local and national diversity panels were previously evaluated. The local diversity panel was characterized in Bambey (Senegal) and in Sadore (Niger) (J. M. Faye et

al., 2019). The national diversity panel was evaluated in Bambey (Senegal) and Nioro (Senegal) (Diack et al., 2020). In the present study, the seeds from 138 lines of the regional diversity panel were phenotyped at both Bambey Research Station (14.42˚ N, 16.28˚ W) and Nioro Research Station (13.45° N, 15.48° W) in Senegal during two rainy seasons in 2020 and 2021. These two stations are representative of the two main agroclimatic zones that divide Senegal and many other countries along the latitude gradient across the continent. The Bambey research station is representative of the semi-arid zone with less than 70 mm of precipitation per year. In the Nioro research station, a subhumid climate is expected with between 700 and 1200 mm of precipitation.

At each location, seeds were sown in mid-July in rows and mounds with spacings of 0.8 m between rows and 0.8 m between mounds within the row. Two additional rows of improved local varieties were used as borders. The experimental design included a 16 x 20 Alpha lattice with two replicates and each elementary plot consisted of a line with 10 mounds per elementary plot. All plants were phenotype for six traits: flowering time, plant height, panicle length, panicle weight, total seed weight and grain yield.

3.5.3 Statistical analysis of phenotypic data

Phenotypic data were analyzed using the R program. Genotype effects were estimated and fitted into a linear regression model to calculate the genotype-adjusted mean value using the *lme4* and *emmeans* package(Bates et al., 2015; Lenth et al., 2025). Phenotypic correlations were calculated using Pearson correlation using the *Performance Analytics* package then visualized with corrplot package (Peterson and Carl 2007 ; Wei et al. 2024). A principal component analysis (PCA) of trait variation was performed to analyze the divergence of ecotypes. Variance components were calculated with the *lme4* package then, Broad-sense heritability (H^2) for each trait was estimated using the following equation:

$$H^2 = \frac{\sigma_G^2}{\left[\sigma_G^2 + \left(\frac{\sigma_{GxR}^2}{R}\right) + \left(\frac{\sigma_{GxY}^2}{Y}\right) + \left(\frac{\sigma_E^2}{RY}\right)\right]},$$

Whereas *G* is genotype, *R* is replication, *Y* is year, and *E* is error (Bates et al., 2025).

3.5.4 Testing fixation of flowering time for core parental lines

Ten landraces were selected from a national diversity panel to represent the major genetic diversity of early- and late-flowering ecotypes in Senegal. These lines were defined in this study as core parental lines, based on prior work by (Diack et al., 2020), who used a heuristic approach to define representative individuals among 392 landraces collected across the country. In their approach, individuals were selected to meet the following criteria: a coincidence rate (%CR) and variable rate (%VR) both greater than 80%, and a mean difference (%MD) and variance difference (%VD) both less than 20%. Using this method, Diack et al. established a national diversity panel of 91 landraces, 60 early-flowering Souna and 31 late-flowering Sanio types. From this set, 14 individuals were identified as the most genetically and phenotypically representative of the major gene pools present nationally. This study selected 10 of those 14 as core parental lines, based on seed availability at the National Center for Agronomic Research (CNRA) in Bambey.

We sowed these seeds at Bambey Research Station (14.42˚ N, 16.28˚ W) in Senegal during two rainy seasons in 2021. The same experimental design as previously described was used for this essay. Four breeding lines were used as controls included the two local improved varieties Souna III and TEMOIN DE SEFA and the two inbred lines ICMB 1777002 and ICMB 177003. At each plot, 12 plants were protected with self bags as soon as they flowered and then self-pollinated. In the hot off season 2022, seeds from self pollinated lines as well as their parental lines and controls were then collected separately and evaluated for the six traits previously cited including flowering time. The distribution and phenotypic variance for flowering time for parental lines of each accession were compared to self-pollinated progenies and controls.

3.5.5 Genotyping data acquisition and SNP discovery and processing

For the genotypic data, we used exome capture for the local diversity panel (J. M. Faye et al., 2019) and genotyping-by-sequencing (GBS) for the national diversity panel (Diack et al., 2020) and regional diversity panel (Kanfany et al., 2020), all generated in previous studies. For GBS data generation, the PstI-MspI enzyme pairs were both used for double digestion before

adding adapters, amplification, and finally sequencing with Illumina platforms (Diack et al., 2020; Kanfany et al., 2020). Reads from the three panels were downloaded from the Short Reads Archive (SRA) and converted to FASTQ format using the fastq-dump() command of the SRA toolbox. Reads were quality analyzed before aligning them separately for each panel to the pearl millet reference genome v2.1 (Ramu et al. 2023) using Burrows-Wheeler Aligner (BWA) v.07 (Li & Durbin, 2010). SNPs were called using Genome Analysis Toolkit 4 (GATK4) v4.2 and raw SNPs were stored in separate vcf files (Auwera & O'Connor, 2020). Biallelic SNPs were selected for the downstream analysis. Selected SNPs were then filtered using the following quality criteria: for quality (QUAL >= 60), for depth coverage (DP > 20), Mapping Quality (MQ) > 40, and Fisher score (FS) < 60. SNPs with Minor Allele Frequency (MAF) > 0.05 and missing rate less than 2% were retained. The final number of SNPs obtained was 233,566 for the local diversity panel, 100,995 for the national diversity panel, and 10,633 for the regional diversity panel.

3.5.6 Genomic analyses

Information about genetic variants in different individuals were stored in a HapMap file then implemented R using the rTassel package (Monier et al., 2022). Neighbor joining trees were constructed using *ape* R package (Paradis & Schliep, 2019). Genome wide association studies (GWAS) were conducted using the GAPIT R package (Lipka et al., 2012). Two statistical models were complementary: the mixed linear model (MLM) and the multi-locus mixed model (MLMM). MLM accounts for relationships among individuals which is included as a kinship matrix. MLMM is an advanced version of the MLM which simultaneously considers the effects of multiple genetic loci by including associated markers as covariates. Manhattan plots and Q-Q plots were generated using the *qqman* R package (Turner, 2014). An *a priori* candidate gene list was defined based on the known flowering genes from cereal crops of foxtail millet (*Setaria italica*), maize (*Zea mays L.*), sorghum (*Sorghum bicolor L.*), and rice (*Oryza sativa L.*) (Table S2). The nucleotide sequence of these genes was BLASTed against the pearl millet reference genome and physical position of candidate ortholog were defined and highlighted in the

Manhattan plot. A SNP position within a 500 kb window with candidate orthologous genes or other annotated genes was defined as colocalization between association signals.

Acknowledgements

We thank our colleague Dr Ghislain Kanfany and the pearl millet breeding team at the Bambey and Nioro research stations for their contribution to this work. We thank Dr. Omar Diack and Dr. Adama Faye for sharing the available phenotypic and genotypic data that we used in part in this study. This study is made possible through funding by the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through grants from American People provided to the United States Agency for International Development (USAID) under cooperative agreement number AID-OAA-A-13-00047 and 7200AA-19LE-00005. The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the US Government.

Chapter 3 References

Auwera, Geraldine A. Van der, and Brian D. O'Connor. 2020. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Inc.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (October):1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, Douglas, Martin Maechler, Ben Bolker [aut, cre, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, et al. 2025. "Lme4: Linear Mixed-Effects Models Using 'Eigen' and S4." https://cran.r-project.org/web/packages/lme4/index.html.

Diack, Oumar, Ghislain Kanfany, Mame Codou Gueye, Ousmane Sy, Amadou Fofana, Hamidou Tall, and Desalegn D. Serba. 2020. "GWAS Unveils Features between Early- and Late-Flowering Pearl Millets," October. https://doi.org/10.21203/rs.3.rs-25381/v3.

Faye, Jacques M., Fanna Maina, Zhenbin Hu, Daniel Fonceka, Ndiaga Cisse, and Geoffrey P. Morris. 2019. "Genomic Signatures of Adaptation to Sahelian and Soudanian Climates in Sorghum Landraces of Senegal." *Ecology and Evolution* 9 (10): 6038–51. https://doi.org/10.1002/ece3.5187.

Franks, Steven J., Sheina Sim, and Arthur E. Weis. 2007. "Rapid Evolution of Flowering Time by an Annual Plant in Response to a Climate Fluctuation." *Proceedings of the National Academy of Sciences* 104 (4): 1278–82. https://doi.org/10.1073/pnas.0608379104.

Grant, Nathan P., John J. Toy, Deanna L. Funnell-Harris, and Scott E. Sattler. 2023. "Deleterious Mutations Predicted in the Sorghum (Sorghum Bicolor) Maturity (Ma) and Dwarf (Dw) Genes from Whole-Genome Resequencing." *Scientific Reports* 13 (1): 16638. https://doi.org/10.1038/s41598-023-42306-8.

Li, Heng, and Richard Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." *Bioinformatics (Oxford, England)* 26 (5): 589–95. https://doi.org/10.1093/bioinformatics/btp698.

Lipka, Alexander E., Feng Tian, Qishan Wang, Jason Peiffer, Meng Li, Peter J. Bradbury, Michael A. Gore, Edward S. Buckler, and Zhiwu Zhang. 2012. "GAPIT: Genome Association and Prediction Integrated Tool." *Bioinformatics (Oxford, England)* 28 (18): 2397–99. https://doi.org/10.1093/bioinformatics/bts444.

Orr, H. Allen. 2009. "Fitness and Its Role in Evolutionary Genetics." *Nature Reviews. Genetics* 10 (8): 531–39. https://doi.org/10.1038/nrg2603.

Peterson, Brian G., and Peter Carl. 2007. "PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis." https://doi.org/10.32614/CRAN.package.PerformanceAnalytics.

Sul, Jae Hoon, Lana S. Martin, and Eleazar Eskin. 2018. "Population Structure in Genetic Studies: Confounding Factors and Mixed Models." *PLoS Genetics* 14 (12): e1007309. https://doi.org/10.1371/journal.pgen.1007309.

Turner, Stephen D. 2014. "Qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots." bioRxiv. https://doi.org/10.1101/005165.

Uffelmann, Emil, Qin Qin Huang, Nchangwi Syntia Munung, Jantina de Vries, Yukinori Okada, Alicia R. Martin, Hilary C. Martin, Tuuli Lappalainen, and Danielle Posthuma. 2021. "Genome-Wide Association Studies." *Nature Reviews Methods Primers* 1 (1): 1–21. https://doi.org/10.1038/s43586-021-00056-9.

Vigouroux, Yves, Cédric Mariac, Stéphane De Mita, Jean-Louis Pham, Bruno Gérard, Issoufou Kapran, Fabrice Sagnard, et al. 2011. "Selection for Earlier Flowering Crop Associated with Climatic Variations in the Sahel." *PLOS ONE* 6 (5): e19563. https://doi.org/10.1371/journal.pone.0019563.

Wang, Hailong, Guanqing Jia, Ning Zhang, Hui Zhi, Lihe Xing, Haoshan Zhang, Yi Sui, et al. 2022. "Domestication-Associated PHYTOCHROME C Is a Flowering Time Repressor and a Key Factor Determining Setaria as a Short-Day Plant." *New Phytologist* 236 (5): 1809–23. https://doi.org/10.1111/nph.18493.

Wei, Taiyun, Viliam Simko, Michael Levy, Yihui Xie, Yan Jin, Jeff Zemla, Moritz Freidank, Jun Cai, and Tomas Protivinsky. 2024. "Corrplot: Visualization of a Correlation Matrix." https://cran.r-project.org/web/packages/corrplot/index.html.

Woods, Daniel P., Thomas S. Ream, Gregory Minevich, Oliver Hobert, and Richard M. Amasino. 2014. "PHYTOCHROME C Is an Essential Light Receptor for Photoperiodic Flowering in the Temperate Grass, Brachypodium Distachyon." *Genetics* 198 (1): 397–408. https://doi.org/10.1534/genetics.114.166785.

Yang, Qin, Zhi Li, Wenqiang Li, Lixia Ku, Chao Wang, Jianrong Ye, Kun Li, et al. 2013. "CACTA-like Transposable Element in ZmCCT Attenuated Photoperiod Sensitivity and Accelerated the Postdomestication Spread of Maize." *Proceedings of the National Academy of Sciences* 110 (42): 16969–74. https://doi.org/10.1073/pnas.1310949110.

Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. 2012. "A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data." *Bioinformatics (Oxford, England)* 28 (24): 3326–28. https://doi.org/10.1093/bioinformatics/bts606.

Chapter 3 Figures and Tables

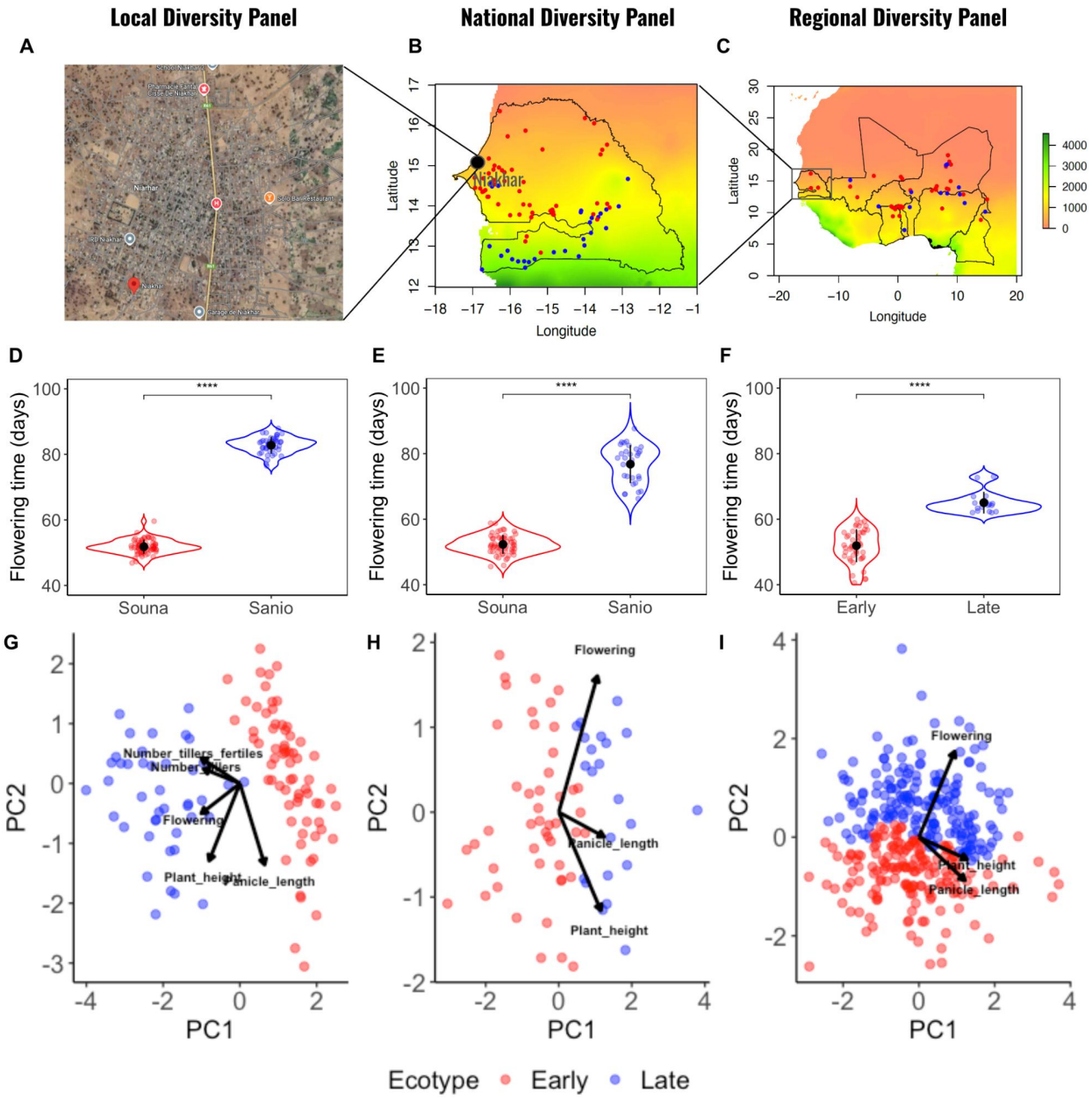


Figure 1. Ecotype divergence within the pearl millet populations that weakens with the widening of the sample scale. Sampling scales of early (red) and late (blue) ecotypes in the Niakhar local diversity panel (A), the Senegal national diversity panel (B), and the West Africa

regional diversity panel. The phenotypic distribution of each panel is presented in D, E, and F, respectively. Ecotypes are classified based on flowering time with statistical significance in each of the three diversity panels. A principal components analysis (PCA) was carried out to analyze the relationships between traits within each diversity panel in the G, H, I panel. A *t*-test was performed to compare the mean between early and late ecotype and the level of statistical significant was coded: $p < 0.001$ = ***, $p < 0.01$ = **, $p < 0.05$ = *.

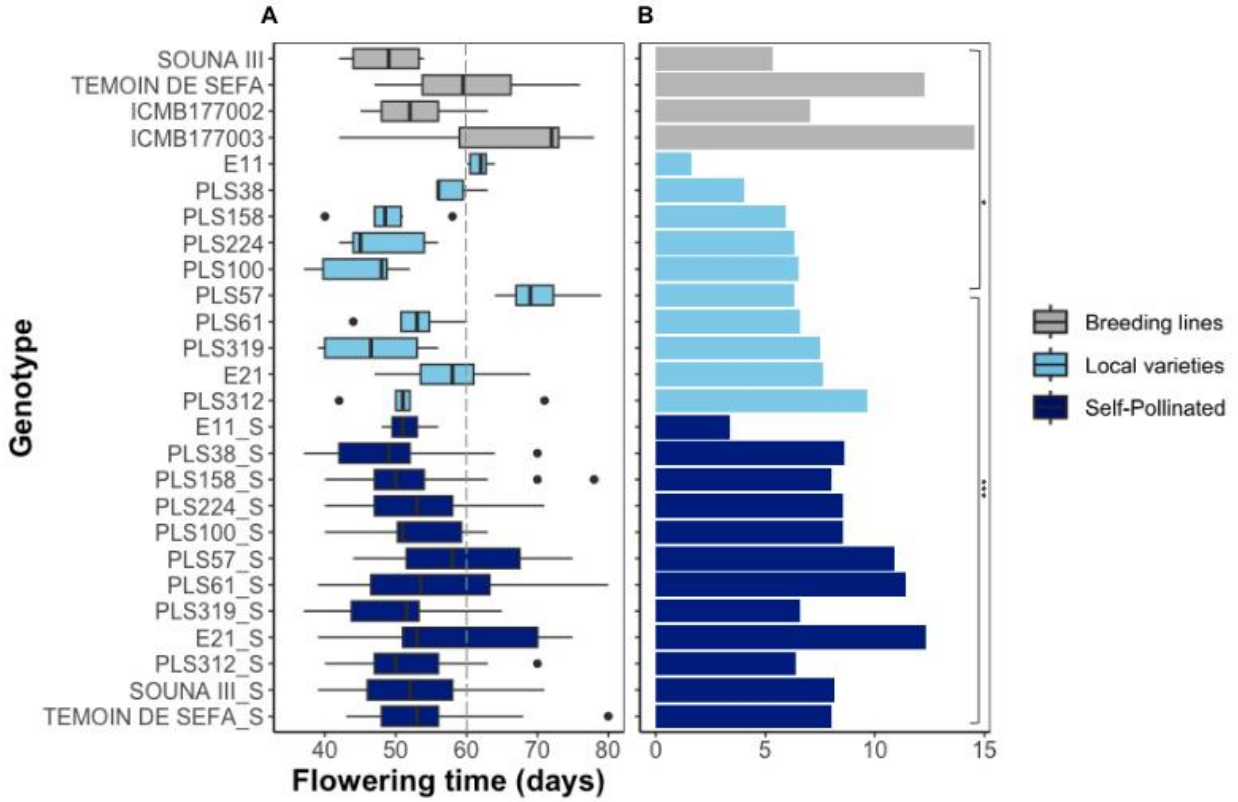


Figure 2. The main flowering loci have not been fixed by natural selection and selection pressure from farmers. Phenotypic distribution (A) and phenotypic variance (B) for flowering time for local varieties that are parental lines (n = 10) from the national diversity panel (light blue), self-pollinated progeny of these local varieties (dark blue), and control breeding lines (gray). The gray dashed lines represented the threshold for flowering time between classified early and late flowering. Two-sample Welch's t-tests were performed to compare the average phenotypic variance between the landrace group and themselves and between the landrace group and breeding lines, respectively. The level of statistical significant was coded: $p < 0.001$ = ***, $p < 0.01$ = **, $p < 0.05$ = *.

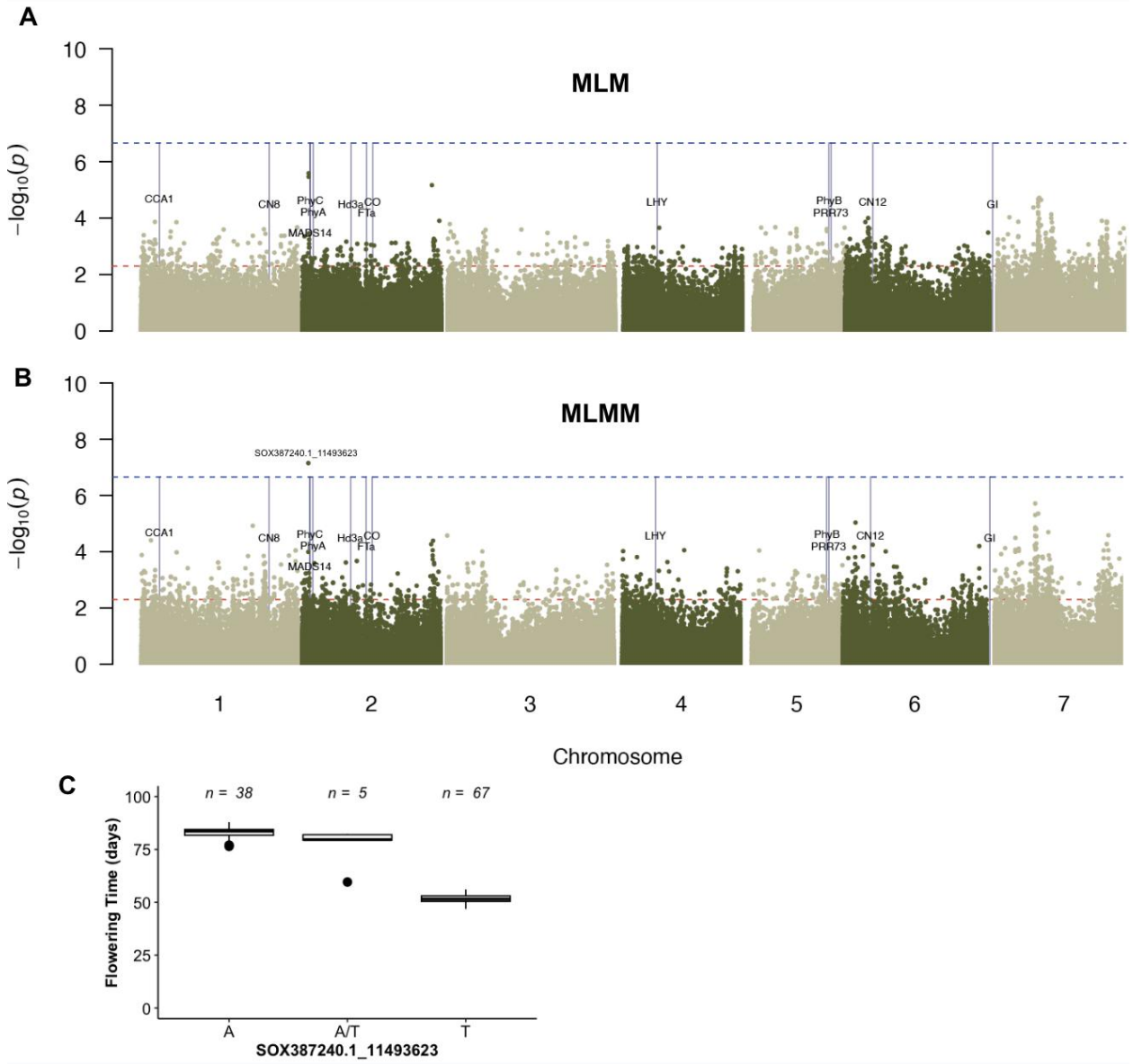


Figure 3. **High-effect variant colocalized with candidate gene Phytochrome (PhyC) in a local diversity panel.** Genome-wide association studies for days to flowering in the Niakhar local diversity panel based on (A) Mixed Linear Model (MLM), and (B) Multiple Locus Mixed Model (MLMM) using 233,566 SNP markers. The blue and red horizontal dotted lines indicate the Bonferroni and FDR significance thresholds at 0.05 and 0.01, respectively. The vertical blue lines represent the position of the candidates flowering time genes. D) Allelic classes at SOX387240.1_11493623 show effect on the days to flowering for the significant SNP based on the MLMM.

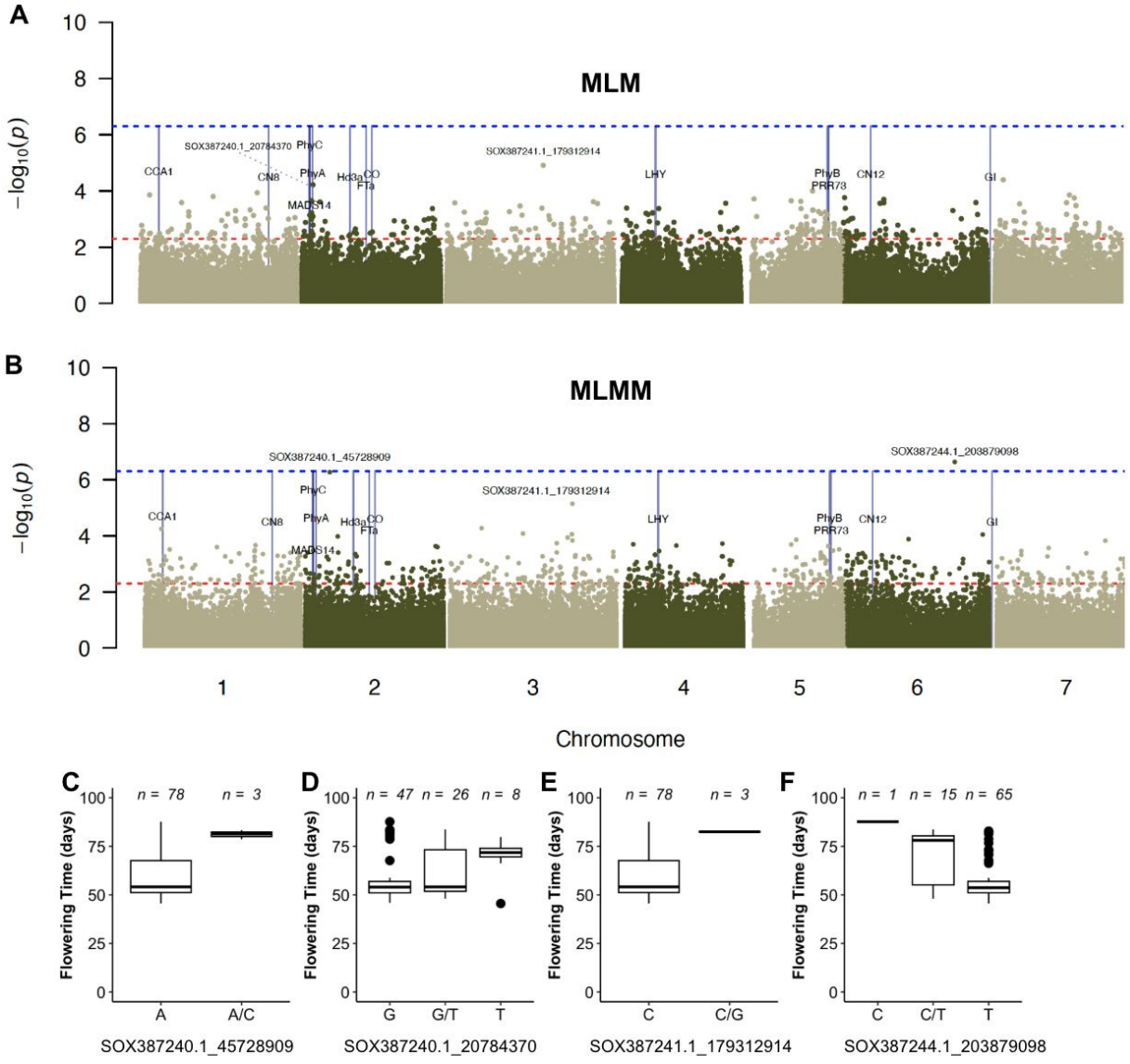


Figure 4. **Oligogenic variation for flowering time in the Senegal national diversity panel.** Genome-wide association studies for days to flowering in the Senegalese national diversity panel using (A) Mixed Linear Model (MLM), and (B) Multiple Locus Mixed Model (MLMM) with 100,995 SNP markers. The blue and red horizontal dotted lines indicate the Bonferroni significance and FDR threshold at 0.05 and 0.01, respectively. Vertical blue lines represent the position of *a priori* candidate flowering time genes. Allelic classes show effect on the days to flowering for two outlier SNPs based on the MLM and MLMM: (C) SOX387240.1_45728909, (D) SOX387240.1_2078370, (E) SOX387241.1_179312914, and (F) SOX387244.1_203879098.

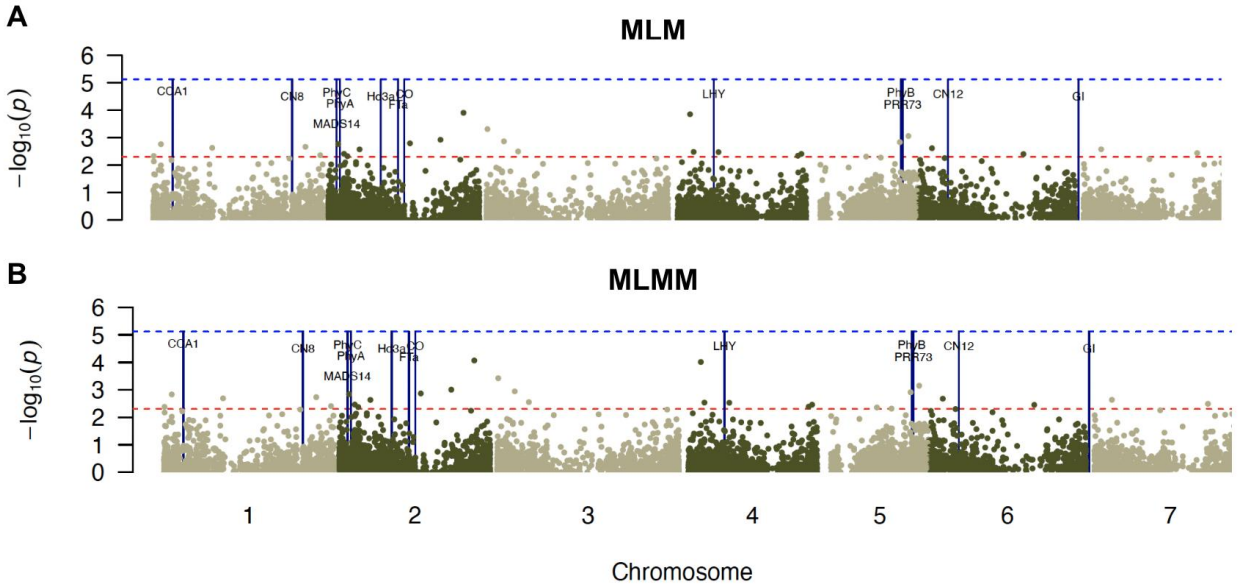


Figure 5. **Oligogenic variation for flowering time in a West African regional diversity panel.** Genome-wide association studies for days to flowering in the regional diversity panel based on A) Mixed Linear Model, and (B) Multiple Locus Mixed Model using 10,633 SNP markers. The horizontal red lines indicate the Bonferroni significance threshold at 0.05. The vertical blue lines represent the position of the candidates flowering time genes.

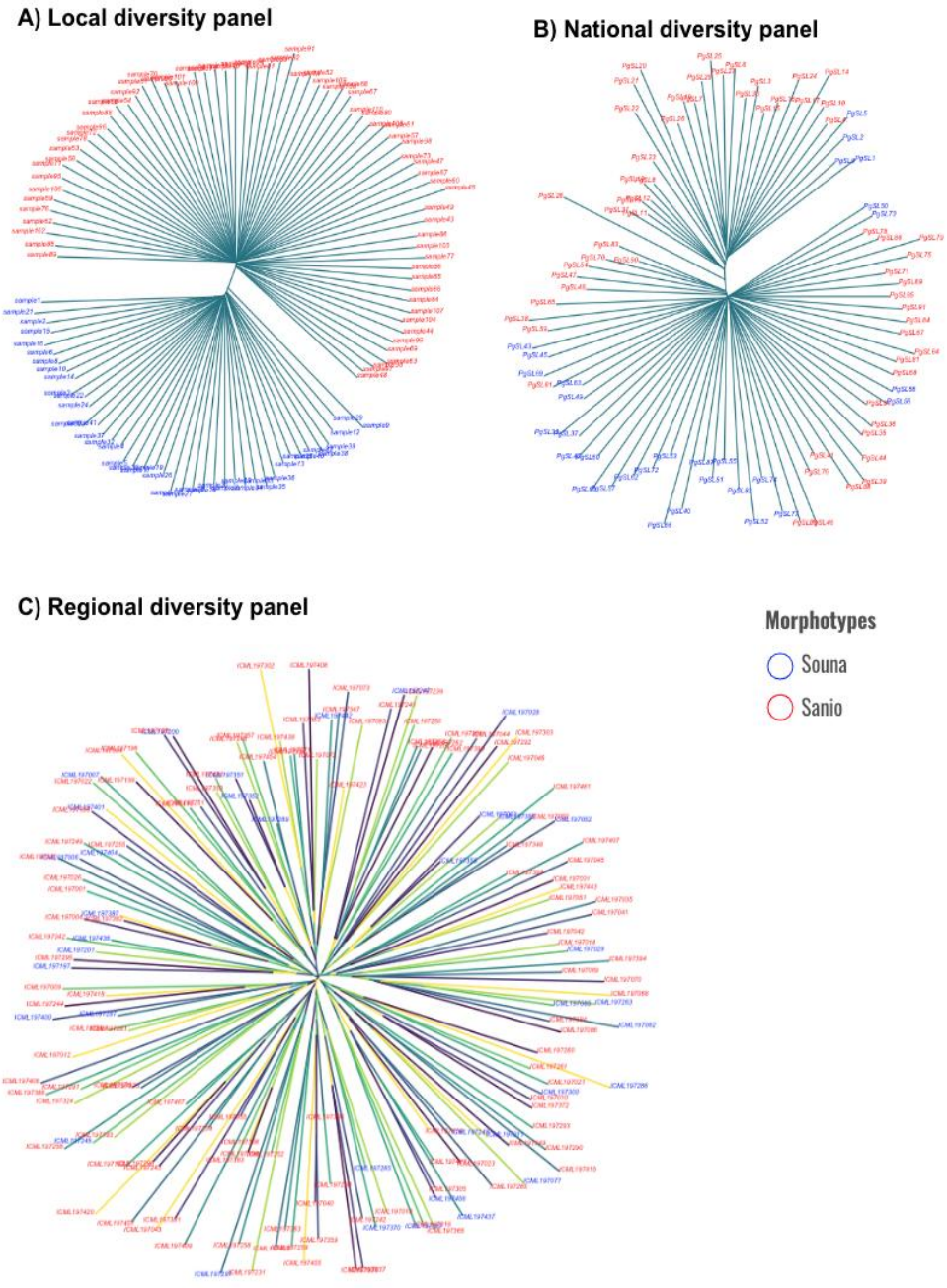


Figure 6. The larger the sampling scale, the more flowering time genetic components are masked in shared gene pools. Neighbor joining trees of the 102 lines from the local diversity panel (A), 90 lines from the national diversity panel (B), and 156 lines from the regional

diversity panel (C). The red and blue colored tip labels correspond to the Souna and Sanio genotypes, respectively. The color of the branches indicates the country of origin of the different lines in West Africa.

Table 1 Broad-sense heritability estimates (H^2) of phenotypic traits across three pearl millet diversity panels.

	Local diversity panel	National diversity panel	Regional diversity panel
Flowering Time	0.99	0.98	0.82
Plant Height	0.93	0.98	0.73
Panicle Length	0.98	0.95	0.96
Total number of tillers	0.77	0.87	–
Tillers Fertile	0.90	–	–
Grain Yield	–	–	0.67

Values represent broad-sense heritability (H^2) estimates (unitless, ranging from 0 to 1). A dash (–) indicates that the trait was not measured or data was not available in the respective panel.

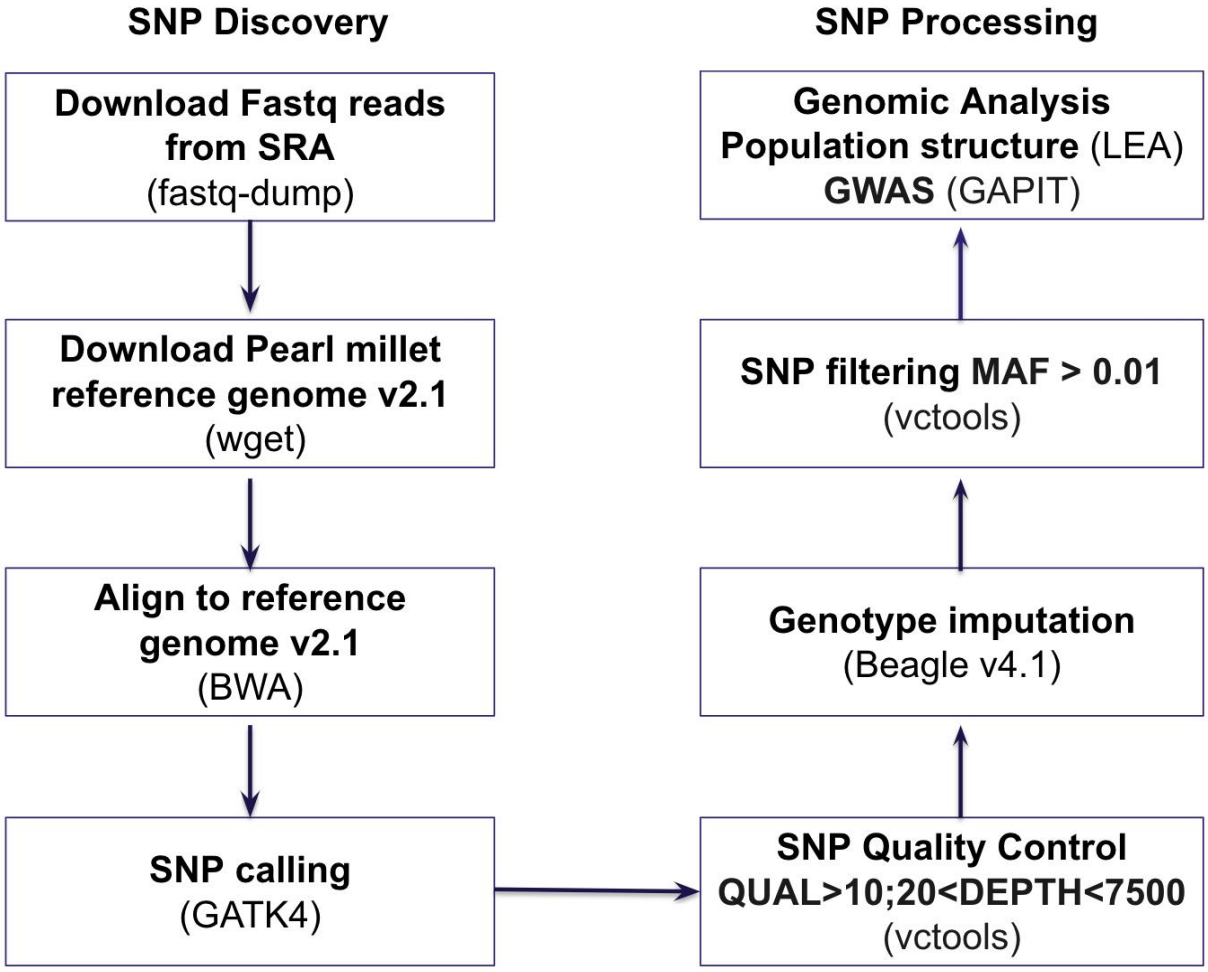


Figure S1. SNP marker discovery and progressing pipeline using genotyping-by-sequencing (GBS)

TableS1. Lines included in the West African regional diversity panel.

ICML197052	SOSATC88-1-1-386	Mali	Landrace
ICML197068	GB8735-1-1-114	Togo	Improved
ICML197069	GB 8735 1-1-282	Togo	Improved
ICML197070	GB8735-1-1-240	Togo	Improved
ICML197071	GB8735-1-1-139	Togo	Improved
ICML197072	GB8735-1-1-066	Togo	Improved
ICML197073	GB8735-1-1-014	Togo	Improved
ICML197077	GB 8735 1-1-300	Togo	Improved
ICML197082	ICTP 8203-1-1-230	Togo	Improved
ICML197083	ICTP 8203-1-1-052	Togo	Improved
ICML197084	ICTP 8203-1-1-159	Togo	Improved
ICML197085	ICTP 8203-1-1-236	Togo	Improved
ICML197086	ICTP 8203-1-1-004	Togo	Improved
ICML197087	ICTP 8203-1-1-109	Togo	Improved
ICML197192	PE00397_B_B_1_1_B_2	Mali	Landrace
ICML197193	PE00349_B 2 1 B	Mali	Landrace
ICML197196	4-2B	Niger	Improved
ICML197197	25-1B	Niger	Improved
ICML197198	47-3B	Niger	Improved
ICML197199	LCICMB1	Niger	
ICML197200	LCICMB6	Niger	
ICML197201	LCICMB7	Niger	

ICML197231	ICMV 167006-B-1-1-B-1	Mali	Landrace
ICML197239	PE02837-B-B-1-1-B	Niger	Landraces
ICML197241	PE00702-B-B-1-1-B	Niger	Landraces
ICML197243	PE00735-B-B-1-1-B	Niger	Landraces
ICML197245	PE02988-B-B-2-1	Niger	Landraces
ICML197246	3/4-ex-Bornu-74-38-2-2-1-1	Nigeria	Landraces
ICML197247	PE01215-B-2	Burkina Faso	Landraces
ICML197248	PE05344-B-1-1-B	Mali	Landraces
ICML197249	SOUNA-159	Senegal	Landraces
ICML197250	PE05434-B-B-1-1-B	Mali	Landraces
ICML197251	PE02695-B-B-1	Niger	Landraces
ICML197252	HKPV-81-1-1-2	Niger	Landraces
ICML197253	ICMV IS 85327-B-1-1	Niger	Landraces
ICML197255	PE02695-B-B-1-1-B	Niger	Landraces
ICML197256	PE01215-B-2-1-B	Burkina Faso	Landraces
ICML197258	ICMV 167006-4-1-1-B	Mali	Landrace
ICML197259	IBL-5-1-1-4	Niger	Landraces
ICML197260	PE02827-B-B-4-1-B	Niger	Landraces
ICML197262	ICTP 8203-Fe-2	Togo	Improved
ICML197263	Souna-109-4-1-1	Senegal	Landraces
ICML197265	IP-2058	Nigeria	Improved cultivar
ICML197281	IP-4927	Senegal	Breeding line

ICML197285	IP-4974	Nigeria	Landrace
ICML197286	IP-5031	Nigeria	Landrace
ICML197287	IP-5131	Niger	Landrace
ICML197288	IP-5253	Niger	Landrace
ICML197289	IP-5272	Niger	Landrace
ICML197290	IP-5438	Niger	Landrace
ICML197291	IP-5441	Niger	Landrace
ICML197292	IP-5560	Niger	Landrace
ICML197293	IP-5695	Nigeria	Breeding line
ICML197294	IP-5713	Nigeria	Landrace
ICML197295	IP-5816	Senegal	Breeding line
ICML197296	IP-5900	Senegal	Landrace
ICML197297	IP-5923	Senegal	Breeding line
ICML197300	IP-6098	Niger	Landrace
ICML197301	IP-6099	Niger	Landrace
ICML197302	IP-6103	Niger	Breeding line
ICML197303	IP-6111	Niger	Landrace
ICML197304	IP-6112	Niger	Landrace
ICML197324	IP-7910	Niger	Breeding line
ICML197342	IP-8426	Nigeria	Landrace
ICML197347	IP-8949	Togo	Breeding line
ICML197348	IP-8972	Togo	Landrace

ICML197350	IP-9282	Togo	Landrace
ICML197351	IP-9301	Togo	Landrace
ICML197352	IP-9347	Ghana	Improved cultivar
ICML197353	IP-9351	Ghana	Landrace
ICML197354	IP-9407	Ghana	Landrace
ICML197355	IP-9446	Ghana	Breeding line
ICML197356	IP-9496	Ghana	Landrace
ICML197357	IP-9651	Nigeria	Landrace
ICML197358	IP-9692	Nigeria	Landrace
ICML197359	IP-9710	Nigeria	Landrace
ICML197363	IP-10085	Mali	Landrace
ICML197365	IP-10343	Nigeria	Breeding line
ICML197370	IP-10539	Senegal	Landrace
ICML197372	IP-10579	Mali	Landrace
ICML197373	IP-10701	Mali	Landrace
ICML197374	IP-10705	Mali	Improved cultivar
ICML197380	IP-11310	Burkina Faso	Landrace
ICML197381	IP-11311	Burkina Faso	Landrace
ICML197382	IP-11346	Burkina Faso	Landrace
ICML197383	IP-11353	Burkina Faso	Landrace
ICML197384	IP-11358	Burkina Faso	Landrace

ICML197386	IP-11577	Burkina Faso	Improved cultivar
ICML197387	IP-11584	Burkina Faso	Landrace
ICML197388	IP-11593	Burkina Faso	Improved cultivar
ICML197394	IP-11984	Nigeria	Landrace
ICML197396	IP-12116	Nigeria	Landrace
ICML197397	IP-12128	Nigeria	Landrace
ICML197399	IP-12298	Nigeria	Breeding line
ICML197400	IP-12322	Nigeria	Landrace
ICML197401	IP-12364	Nigeria	Landrace
ICML197404	IP-12925	Ghana	Landrace
ICML197406	IP-13016	Mali	Breeding line
ICML197407	IP-13149	Niger	Landrace
ICML197408	IP-13154	Niger	Landrace
ICML197409	IP-13180	Nigeria	Landrace
ICML197414	IP-13817	Burkina Faso	Breeding line
ICML197415	IP-13840	Burkina Faso	Landrace
ICML197423	IP-15533	Burkina Faso	Breeding line
ICML197424	IP-15553	Burkina Faso	Landrace
ICML197427	IP-15917	Togo	Landrace
ICML197436	IP-17554	Togo	Landrace
ICML197437	IP-17611	Togo	Landrace

ICML197442	IP-18157	Mali	Landrace
ICML197443	IP-18168	Mali	Landrace
ICML197454	IP-19584	Niger	Landrace
ICML197455	IP-19612	Niger	Landrace
ICML197456	IP-19613	Niger	Landrace
ICML197457	IP-19626	Niger	Landrace
ICML197458	IP-21020	Nigeria	Landrace
ICML197461	IP-21517	Niger	Landrace
ICML197467	IP-22494	Niger	Landrace

Information on the pedigree, country of origin, and biological status of each line.

Chapter 4: Gene regulatory network for floral transition underlying ecotypic divergence in pearl millet

4.1 Summary

Floral transition is tightly controlled by plants in response to endogenous and environmental cues, as its timing affects adaptation and yield in crops. Adaptation to drylands requires earlier flowering millet varieties that escape terminal drought events. In pearl millet (*Cenchrus americanus*), a major ecotypic divergence is driven by variation in genetic factors belonging to a gene regulatory network that is highly conserved across plant species. However, little is known about the evolution of this network in pearl millet. This study aims to investigate key flowering genes underlying differential flowering between early Souna and late Sanio genotypes. We used inbred lines adapted to the West African ecosystem to study their differential expression between the early Souna and the late Sanio genotypes at vegetative and flowering stages. Phylogeny analyses detected orthologs and variation in sequences in the pearl millet reference genome (Tift23D2B1-P1-P5), suggesting conserved genes function and specialization in pearl millet. Gene expression study of key flowering regulators reveals significant differential expression between early and late genotypes ($p < 0.001$), indicating that the floral transition is triggered by different levels of florigens. *CaFTa* and *CaPhyA* repress the expression level of activator genes (*CaHd3a* and *CaFTb*) to monitor the GI-dependent pathway and delay flowering. By contrast, high expression of *CaHd3a* is associated with early flowering in early Souna and *CaFTb* ($p < 0.001$) triggers flowering in late Sanio at later stages. The results also revealed different coexpression patterns between photoperiod signals and floral integrators. Therefore, we propose a hypothetical model of pearl millet flowering regulation where the photoperiod activates flowering through the interaction between floral activators/repression switch to monitor the GI-dependent pathway. Validation of these hypotheses would lay the foundation for understanding the regulation of millet floral transition to identify key genes whose expression should be monitored to achieve an optimal flowering phenotype.

4.2 Introduction

Photoperiodic flowering is one of the main mechanisms allowing crops to migrate and adapt to different environmental gradients (Bradshaw & Holzapfel, 2017). From its center of domestication around northeastern Mali, pearl millet spread throughout the world and survived very harsh temperatures and precipitation deficits across different latitude gradients (Manning et al., 2011). At higher latitudes, the day is longer and "Souna" ecotypes flower early, 50 to 60 days after planting, so that they can complete their cycle quickly, escape unfavorable temperatures and terminal droughts, and maximize grain production (Diack et al., 2017). In wetter southern regions, days are shorter, "Sanio" ecotypes flower between 80 and 110 days after planting, to maximize biomass production (Diack et al., 2017). Thus, the flowering period, by determining the length of the cycle, influences the occurrence and severity of biotic and abiotic stresses such as drought (Kazan & Lyons, 2016). Flowering is therefore a particularly important trait whose understanding of the control mechanisms is of great interest for successful breeding and productive agriculture.

Floral induction is triggered by the interaction between endogenous and environmental factors (Chen & Li, 2022; Song et al., 2013). Endogenous signals indicate the extent to which the plant has stored sufficient resources to change developmental phases (Song et al., 2013). In addition, environmental factors signal favorable conditions for optimal seed production (Song et al., 2013). This interaction, commonly known as the external coincidence model, was discovered in *Arabidopsis thaliana* (Pittendrigh & Minis, 1964). According to the model, Flowering Locus T (FT) proteins accumulate up to a certain threshold and then move from the leaves, where they are produced, to the meristem to induce flowering under long-day conditions.The induction of *FT* genes itself is positively regulated by the CONSTANS (*CO*) protein, which in turn is negatively regulated by a family of proteins called Cycling DOF Factors (*CDF*) (Pittendrigh & Minis, 1964; X. Wang et al., 2021). At the end of the day, the FLAVIN-BINDING KELCH REPEAT F-BOX 1 (*FKF1*) and GIGANTEA (*GI*) complex stabilized by light degrades CDF proteins, thereby facilitating the expression of *CO* (Pittendrigh & Minis, 1964; X. Wang et al., 2021).

These upstream components of the flowering pathway particularly CO, FT, and GI are transcriptionally regulated in leaf tissue, where environmental cues such as photoperiod are first perceived and integrated. This tissue-specific expression is thus a key component of floral timing regulation, justifying the use of leaf samples in expression studies aiming to capture the regulatory dynamics during the vegetative and transition phases. Further studies have revealed more than a hundred flowering genes involved in this pattern that were found to be highly conserved at the sequential and functional level across plant families (Blümel et al., 2015; Gaudinier & Blackman, 2020).

A divergent regulatory network was found in the flowering pathway conserved among crop species. Orthologs of Arabidopsis *FT*, *CO*, *GI* and others key regulatory genes have been identified in crop species, including major cereals such as rice, maize, sorghum, foxtail millet, wheat, and barley (Y. Li & Xu, 2017; H. Wang et al., 2022). In these crops, the function and regulatory relationship between CO and FT are conserved; however, CO expression is induced by different day-length conditions: short days for rice, maize, and sorghum, and long days for wheat and barley (Y. Li & Xu, 2017). Additionally, functional divergence has been observed in key environmental signals. For example, the circadian clock-associated transcription factor pseudo-response regulator 37 (PRR37) in sorghum acts as a repressor of florigen by blocking EARLY HEADING DATE 1 (Ehd1) expression during long days (Murphy et al., 2011). By contrast, PRR37 induces the major corn florigen, ZEA CENTRORADIALIS 8 (*ZCN8*), by inhibiting its repressor ZmCCT during long days (Y. Li & Xu, 2017). This functional divergence may result from non-synonymous mutations affecting protein function or regulatory variation in gene expression. Among these mutations, those beneficial for adaptation to specific agro-ecological conditions are more likely to be fixed (Greenbaum et al., 2016). Although conserved function between orthologous genes has been observed across distant species, there is growing evidence that evolutionary processes can lead to functional divergence even in closely related species.

Although the flowering regulation pathway has been well described for many cereal crops, little is known about pearl millet. Two homologous genes, Heading date 3a (*Hd3a*) and D*warf8* genes, were under selection during domestication (Lakis et al., 2012; Mariac et al.,

2006). High phenotypic variation for flowering was observed, particularly in West African landraces, with high heritability. For example, the Phytochrome C (*PHYC*) gene polymorphism was associated with phenotypic variation in flowering time, spike length, and stem diameter (Saïdou et al., 2009). A significant increase in the early- flowering phenotypes was also associated with signatures of selection at the *PHYC* locus between 1976 and 2003 in response to drought episodes in Niger (Vigouroux et al., 2011). Additional association mapping studies identified QTLs for flowering time, with few co-localizations to known flowering time genes in other grasses such as Heading date 16 (*Hd16*) and *PhyC* (Diack et al., 2020; Faye et al., 2022; Kumar et al., 2017).These findings suggest that flowering time variation in pearl millet is genetically controlled by variation in conserved regulatory genes. However, no integrated regulatory model has yet been proposed for this species.

Here, we study the gene regulatory network underlying the major ecotypic divergence in flowering time between Souna and Sanio pearl millet. We identified 16 candidate homologs of core flowering regulator genes of other cereal crops in the millet reference genome. We performed phylogenetic and protein pattern analysis to test whether the candidate genes had conserved the genetic structure of their closely related species. We analyzed the differential expression of these genes between inbred lines of early- and late-flowering ecotypes from the West African Millet Inbred Germplasm Association Panel (PMIGAP) during the vegetative and flowering stages to test their putative role in regulating floral transition. Importantly, since critical regulators of flowering such as FT, CO, and GI are expressed in leaf tissue where they perceive environmental cues and initiate the flowering cascade, our expression analysis was performed using leaf samples to best capture the upstream regulatory events. We then propose a flowering regulation model based on homology and gene coexpression patterns in early- and late -flowering ecotypes across developmental stages.

4.3 Results

4.3.1 Genomic analyses reveal conservation and loss of flowering genes in pearl millet

To identify candidate flowering time genes in pearl millet, a list of 22 core flowering regulators was queried against the pearl millet reference genome *(v2.1) using BLAST (Table S2).*

Homologs for 20 of the 22 genes were identified. Notably, *PRR37* and *GRAIN NUMBER, PLANT HEIGHT AND HEADING DATE 7 (GHd7)*, two major floral repressors under long-day conditions in rice, sorghum (Ma1, Ma6), and maize, were absent from the pearl millet genome (Murphy et al., 2011; Zheng et al., 2019). Gene positions are detailed in Table S2 and visualized in Figure S1. In cases where multiple hits were retrieved, the match with the highest identity percentage, bit score, and lowest E-value was retained for further analysis.

Interestingly, some candidate florigen genes colocalized at the same chromosomal positions in pearl millet. For example, Heading date 3a (*Hd3a*) and RICE FLOWERING LOCUS T 1 (*RFT1*), typically mapped on different locations on chromosome 6 in rice, were both found on chromosome 2 in pearl millet. These genes act at different developmental stages in rice: Hd3a initiates early floral transition, while *RFT1* functions at later stages. Similarly, foxtail millet florigens *SiFTa* and *SiFTb*, which have opposing effects on flowering, also mapped to the same region in the pearl millet genome. Given their clustering and functional divergence in other species, the top five hits for florigen genes were retained for phylogenetic analysis. In total, 20 candidate genes including five florigen genes were used in subsequent analyses.

4.3.2 High protein similarity suggests conserved function for some, but not all, flowering genes

An a priori regulatory model was constructed based on known functions of candidate gene orthologs in related grass species (Figure 1D). To assess whether candidate genes retained their ancestral functions in pearl millet, we examined protein sequence similarity, conserved domains, and phylogenetic relationships. Phylogenetic and motif analyses showed strong sequence conservation for several genes, including *CaGI*, *CaEhd1*, *CaHd6*, *CaCCA1*, *CaLHY*, and the phytochrome and pseudo-response regulator families (Figures 2 and S2). For instance, CaGI grouped within the same clade as its orthologs from *Setaria*, *Zea*, and *Sorghum* (*SiGI*, *ZmGI*, *SbGI*), indicating functional conservation (Figure 2A).

The *CaCO* gene, while containing an additional N-terminal sequence not present in orthologs, maintained conservation in key functional domains: the tandem B-boxes (protein-protein interaction) and the C-terminal CCT domain (light signaling) (Figure S2). Despite

phylogenetic separation from monocots and dicots, the preservation of these domains suggests that *CaCO* has likely retained its regulatory role. Unexpected clustering was observed for *CaHd3a*. Rather than grouping with *OsHd3a* or *OsRFT1*, its closest nucleotide matches, it clustered with SiFTa and *SiFTb*, which are known to have divergent flowering functions (Figure 2B), suggesting potential functional divergence or novel regulation in pearl millet.

4.3.3 Different florigens tigger floral transition in Souna and Sanio genotypes

To determine whether differences in gene expression drive ecotypic divergence, we conducted qRT-PCR on 16 candidate genes at both vegetative and flowering stages in early-flowering (Souna) and late-flowering (Sanio) genotypes (Figure S3). In Souna genotypes, *CaCO* was more highly expressed at the vegetative stage, while *CaHd3a* showed significantly higher expression at the flowering stage (Figures 3A–C). Although *CaHd3a* was also induced at the flowering stage in Sanio genotypes, its expression remained lower than in Souna. Conversely, *CaFTa* and *CaFTb* were more strongly expressed in Sanio, particularly at the flowering stage (Figure 3D), suggesting they play a larger role in floral transition in these late-flowering genotypes.

Notably, *CaCN8*, another florigen candidate, exhibited higher expression at the vegetative stage in Sanio, potentially contributing to delayed flowering. Floral integrators *CaCO*, *CaGI*, and *CaEhd1* displayed consistently higher expression in Souna (Figure 3A). Photoperiod-related genes exhibited more mixed patterns: *CaPhyB*, *CaPhyC*, and *CaLHY* were more expressed in Sanio at the vegetative stage, while *CaPhyA* was more expressed in Souna (Figures 3A and 3D). These patterns suggest that early induction of *CaHd3a* may initiate floral transition in Souna, while *CaFTa*, *CaFTb*, and *CaCN8* likely drive floral transition in Sanio through alternative pathways.

4.3.4 Interaction between photoperiod signals and florigen genes drives ecotypic divergence

To understand how gene interactions differ between early and late ecotypes, we analyzed gene expression correlations by genotype and developmental stage (Figures 3B–F). At the vegetative stage, *CaPRR95* positively correlated with *CaGI*, while *CaPhyA* showed a negative correlation with floral activators, especially *CaHd6*, in Souna (Figure 3B). In contrast, Sanio

genotypes exhibited strong positive correlations between florigens and *CaPhyA*, *CaPhyC*, and *CaPRR73* (Figure 3E), suggesting an alternative pathway operating under long-day conditions. At the flowering stage, Souna genotypes showed robust associations between *CaPRR73*, *CaPRR95*, and floral integrators. However, *CaFTa* was negatively correlated with CaCO, indicating a distinct regulatory behavior (Figure 3C). In Sanio, floral gene correlations were weaker and more selective: *CaPhyC* was positively correlated with *CaCN8*, *CaHd3a,* and *CaGI*, while *CaPhyB* correlated with *CaFTa* and *CaPhyA* showed a negative correlation with *CaHd6* (Figure 3F). These results support the hypothesis that distinct photoperiodic signaling pathways regulate floral transition in Souna and Sanio.

4.3.5 Photoperiodic signals via CaGI-depend pathway trigger flowering in Souna

To test whether flowering is regulated by the same upstream pathways in both ecotypes, we constructed gene co-expression networks (Figure 4). In Souna, a highly interconnected network emerged at the vegetative stage, with central nodes around *CaGI* and *CaHd3a*, while *CaPhyC* appeared isolated (Figure 4A). This structure suggested early-stage regulation through a *CaGI*-dependent pathway. At flowering, the network became denser, with strong links among *CaPRR95*, *CaGI*, *CaHd6*, and most florigens except *CaFTa*, which remained disconnected and potentially disruptive to the core module (Figure 4C). Notably, phytochromes A, B, and C, and *CaMADS14*, were peripheral in this network.

In Sanio, the vegetative-stage network showed sparse interactions and two disconnected clusters (Figure 4B), suggesting fragmented signaling. At flowering, the strongest interactions revolved around CaPhyA, CaPhyC, and CaPRR73, with florigens largely decoupled from CaGI (Figure 4D). Together, these results indicate that Souna flowering is triggered through a canonical CaGI-dependent photoperiodic pathway, while Sanio flowering relies on alternative networks involving CaPhyA/C and PRRs. This reflects a divergence in regulatory mechanisms underpinning early vs. late flowering ecotypes.

4.4 Discussion

4.4.1 Distinct temporal CaHd3a and CaFTb drives ecotype divergence in flowering time

This study provides a first mechanistic framework for flowering time regulation in *Pennisetum glaucum* (pearl millet), highlighting key differences between early-flowering (Souna) and late-flowering (Sanio) ecotypes. We identified three major FT-like genes *CaHd3a*, *CaFTa*, and *CaFTb* orthologous to *OsHd3a*, *SiFTa*, and *SiFTb*, that exhibit divergent spatial and temporal expression patterns. In Souna, rapid activation of the CaGI-dependent pathway around 60 days after sowing leads to a burst of CaHd3a expression sufficient to initiate flowering. In contrast, in Sanio, this pathway remains repressed until ~90 days, and flowering occurs only after cumulative expression of CaFTb (Figure 3, 4). This suggests *CaFTb* acts as a compensatory florigen under long-day (LD) conditions in late ecotypes.

This sequential and potentially redundant florigen activation is reminiscent of *Hd3a* and *RFT1* activity in rice (Komiya et al., 2008), and illustrates a form of pathway buffering that allows phenotypic plasticity under variable environments. Interestingly, CaFTa may act antagonistically especially in Sanio by suppressing *CaHd3a* expression. This mirrors inhibitory roles of *TSF* in Arabidopsis and *SiFTb* in *Setaria* (H. Wang et al., 2022; Wolabu et al., 2016; Zheng et al., 2019), reinforcing that FT-like genes can undergo neofunctionalization, acquiring opposite regulatory roles in different genetic or ecological contexts.Phytochromes as key modulators of divergent flowering pathways

4.4.2 Phytochromes as Gatekeepers of photoperiod sensitivity

Our analysis suggests that *CaPhyC*, along with *CaPhyB* and *CaPhyA*, contributes to the divergence in photoperiod sensitivity between the two ecotypes. *CaPhyB* shows higher expression in Sanio at the vegetative stage and likely acts as a flowering repressor by destabilizing *CaCO*, thereby inhibiting the *CaGI*-dependent pathway (Figures 3D, 3F). In support, *AtPHYB* is known to promote CO degradation under red light and delay flowering (Mockler et al., 2003), while *SbPHYB* (Ma3 locus) suppresses flowering under LD in sorghum (S. Yang et al., 2014). Of particular interest is *CaPhyC*, whose expression is delayed in Sanio and disconnected from *CaPhyB*, yet highly correlated with *CaPhyA* (Figure 3E). This pattern

suggests a potential novel interaction where CaPhyA may stabilize CaPhyC, perhaps through heterodimerization, a mechanism previously reported for PHYB/PHYC in rice (Xie et al., 2014). This would constitute a novel function for PHYTOCHROME A in pearl millet, possibly acquired during adaptation to arid Sahelian environments.

Our phylogenetic analysis supports a conserved origin for *CaPhyC*, but with signs of domain loss at the C-terminal signaling region compared to *OsPhyC*, possibly due to subfunctionalization during speciation within Panicoideae (Figure S2). Despite this, the photosensory domains remain intact, indicating conserved light sensing, but potential divergence in downstream signaling (J. Li et al., 2011; Monte et al., 2003). Importantly, CaPhyC appears to delay the floral transition in Sanio by suppressing early activation of CaHd3a (Figures 3, 4). This behavior resembles the role of PhyC in *Arabidopsis*, *rice*, *foxtail millet*, *maize*, and *sorghum*, where mutations promote early flowering(Su et al., 2016; Sun et al., 2022; H. Wang et al., 2022). However, the opposite effect is observed in *wheat* and *Brachypodium*, where *PhyC* mutants delay flowering (Kippes et al., 2020). This supports the broader observation that orthologous genes can acquire opposing roles across species, reflecting both evolutionary constraint and plasticity.

4.4.3 Gene Co-expression Network Reveals Divergence in Regulatory Hubs

Our network analysis highlights differences in gene regulation between ecotypes. In Souna, early flowering is linked to a highly connected gene network centered on *CaGI*, *CaHd3a*, and *CaPRR95*, which follows a typical photoperiodic flowering model (Figure 4A). Similar central hubs have been seen in other species where flowering is strongly controlled by day length (Davila-Velderrain et al., 2016; Ferreira et al., 2018).

In contrast, the Sanio ecotype shows a more fragmented network that changes depending on the developmental stage. In this case, *CaPhyC*, *CaPhyA*, and *CaPRR73* appear to play more important roles in regulating flowering (Figure 4D). This shift in regulatory control suggests that different genetic pathways are active in each ecotype, which is also supported by previous gene network studies in plants (X.-F. Yang et al., 2024).

The networks also reveal genes that may have different functions in each ecotype. For example, *CaCN8*, which is a possible florigen gene, promotes flowering in Souna, but in Sanio it

shows a negative correlation with flowering genes (Figure 5). This kind of opposite behavior has been seen in other crops like rice, where changes in the DNA sequence of florigen genes (like *Hd3a*) affect their function under different conditions (Tamaki et al., 2007; Vicentini et al., 2023). It is possible that *CaCN8* carries mutations in Sanio that change how it works or how it interacts with other genes.

4.4.4 A first regulatory model for flowering time regulation: implication for breeding and drought adaptation

This study proposes the first hypothetical model for flowering time regulation in *Cenchrus americanus*, revealing ecotype-specific regulatory networks that may underlie adaptation to contrasting environments. In early-flowering Souna, flowering is initiated by strong activation of the CaGI–CaHd3a module, suggesting a photoperiod-sensitive pathway that enables rapid floral transition. In contrast, late-flowering Sanio appears to bypass this route, relying instead on CaFTb, which may act as a compensatory florigen under long-day conditions.
A key feature of this model is the distinct role of phytochromes. In Sanio, CaPhyC and CaPhyA may suppress early flowering by delaying floral activator activity at the vegetative stage. This inhibition is lifted later, allowing flowering to proceed without strong central input from CaGI. Previous studies have linked nonsynonymous SNPs in CaPhyC with flowering time variation in both Senegalese and Nigerien germplasm (Diack et al., 2020; Faye et al., 2022; Saïdou et al., 2009), highlighting its value as a marker for breeding programs.

Another important finding is the ecotype-specific behavior of *CaCN8*, a florigenic candidate. It acts as a flowering promoter in Souna but shows negative correlations with floral genes in Sanio. This kind of antagonistic effect has been seen in rice, where functional changes in Hd3a due to single nucleotide polymorphisms alter flowering behavior depending on environmental context (Itoh et al., 2010; Komiya et al., 2008). These findings suggest that CaCN8 might carry adaptive mutations relevant for drought escape.

Given the lack of transformation protocols in pearl millet, functional validation will require alternative strategies, such as bulk segregant analysis, association mapping, or transient assays in heterologous systems. Looking ahead, pangenome approaches could provide deeper

insight into structural and presence–absence variation in flowering genes. FT-like genes, including *CaHd3a* and *CaCN8*, are known to undergo duplication and subfunctionalization (H. Wang et al., 2022), making them strong targets for allele mining. Together, these results lay the foundation for molecular breeding tools, such as qPCR- or KASP-based markers, to select for early-flowering, drought-adaptive lines in pearl millet across West Africa.

4.5 Conclusion

This study reveals distinct genetic and regulatory mechanisms controlling flowering time in pearl millet ecotypes, with early flowering driven by rapid activation of CaHd3a and late flowering regulated by delayed expression of CaFTb and phytochrome-mediated repression. The differential expression and co-expression patterns of key flowering genes highlight adaptive strategies to photoperiod and drought conditions in West African environments. These insights provide a valuable foundation for breeding programs aiming to develop early-flowering, drought -tolerant millet varieties, essential for improving crop resilience and food security in drylands. Future research focusing on functional validation and allele mining will enhance our understanding of flowering regulation and enable precision breeding. Ultimately, this work advances efforts to tailor pearl millet to changing climates and diverse agricultural needs.

4.6 Materials and methods

4.6.1 Identification of candidate flowering time genes in pearl millet

A set of core flowering regulatory genes conserved across cereal species was compiled from the literature (Table S1). The coding sequences (CDS) and full-length protein sequences of these genes from both monocot and dicot plant species were retrieved from Phytozome and NCBI databases. To identify their homologs in pearl millet (*Pennisetum glaucum*), a local BLASTN database was constructed using *makeblastdb* from the BLAST+ 2.12.0 suite. The CDS of core genes from closely related grass species—*Setaria italica*, *Sorghum bicolor*, *Zea mays*, and *Oryza sativa*—were queried against the pearl millet reference genome (P. glaucum v2.1). For each core gene, the best hit was selected based on the lowest E-value, highest percentage

identity (pident), and highest bit score. Chromosomal positions of each candidate match were compared against the pearl millet genome database (GIGADB: http://gigadb.org/dataset/100192). Gene IDs corresponding to each core gene ortholog were identified, and their CDS and protein sequences were extracted (Table S2; Figure S3). The chromosomal positions of candidate genes were visualized using PhenoGram software (https://ritchielab.org/software/phenogram-downloads).

4.6.2 Multisequence alignment, phylogenetic analysis, and conserved protein motif identification

The retrieved nucleotide sequences of candidate genes and their homologs from both monocots and dicots were aligned using MUSCLE within the Molecular Evolutionary Genetics Analysis software (MEGA11) with default parameters. Phylogenetic trees were constructed using the maximum likelihood method with 1,000 bootstrap replicates. The corresponding full-length protein sequences were analyzed for conserved motifs using the Multiple EM for Motif Elicitation (MEME) tool (https://meme-suite.org/meme/tools/meme) in classic discovery mode, setting the maximum number of motifs to 20 and keeping all other parameters at default. Motif visualization and annotation were performed using TBtools (v1.098728). Protein identity percentages between pearl millet proteins and their homologs in other species were calculated using NCBI BLASTp.

4.6.3 Plant materials and growth conditions

Three early-flowering (Souna) lines—ICML 197281 (IP-4927), ICML 197352 (IP-5695), and ICML 197347 (IP-8949)—and three late-flowering (Sanio) lines—ICML 197297 (IP-5923), ICML 197373 (IP-10701), and ICML 197458 (IP-21020)—were selected from the Pearl Millet Inbred Germplasm Association Panel (PMiGAP) based on ecotype classification and seed availability (Diack et al., 2020; Table S6; Figure 1A–C). These lines were selected based on ecotype criteria and seed availability (Figure 1 A-C). Seeds were provided by the Centre National de Recherches Agronomiques (CNRA/ISRA) and sown in July 2022 under long-day conditions in a greenhouse at the National Laboratory for Plant Production Research (LRNPV/ISRA). A completely randomized block design was used with three replicates per

genotype. For each replicate, six plants per genotype were grown and used for both flowering time measurement and gene expression analysis.

4.6.4 Primers design and gene expression analysis

For each plant, the entire uppermost leaf was harvested at both vegetative and reproductive stages for RNA extraction. Samples were flash-frozen in liquid nitrogen and ground to a fine powder. Total RNA was extracted using TRIzol™ Reagent following the manufacturer's instructions. RNA integrity was verified on a 1.2% agarose gel, and RNA concentration and purity were assessed using a NanoDrop spectrophotometer. For each sample, 1 µg of total RNA was reverse-transcribed into cDNA using the GoScript™ Reverse Transcriptase system (Promega) following the manufacturer's protocol. Resulting cDNA was diluted 5-fold prior to use in quantitative PCR.

Primers for candidate flowering time genes were designed using Primer3 software and are listed in Table S5. qRT-PCR was performed using the StepOnePlus™ Real-Time PCR System under SYBR Green settings with GoTaq® qPCR Master Mix (Promega), which includes BRYT Green® dye. Specificity of primer amplification was confirmed via melting curve analysis (Figure Sx). Two housekeeping genes, Actin and TIP41, were used for normalization. Each reaction included three biological and two technical replicates. A single reference sample was included on all plates for normalization using StepOne v2.3 software (Applied Biosystems, Grand Island, NY, USA). For cross-run normalization, qbase+ v3.4 (Hellemans et al., 2007) was used with the following settings: single amplification efficiency applied across targets, normalization based on global mean, and average scaling. Results were exported as calibrated normalized relative quantities (CNRQ values). Expression values represent the mean of the three biological replicates and were normalized to CaTIP41 expression.

4.6.5 Gene co-expression network analysis

Normalized expression data were used to calculate pairwise correlation coefficients and p -values using the *rcorr()* function from the *Hmisc* R package (Harrell Jr, 2003). The resulting correlation matrix was visualized using the *corrplot* package (Wei et al., 2024), with significance thresholds set at $p < 0.01$ and a 95% confidence level. Separate correlation plots were generated

for early- and late-flowering genotypes, as well as for the vegetative and flowering stages. Gene interaction networks were constructed using the *qgraph* R package (Epskamp et al., 2023). Only statistically significant edges ($p < 0.05$) were included to represent co-expression links between genes.

Acknowledgements

We thank Dr Adama Faye and the team at the National Laboratory for Research on Plant Production (LNRPV) in Dakar, Senegal for access to all the equipment and infrastructure we used to conduct this study. We thank our colleague Dr Ghislain Kanfany and Dr Omar Diack and through them the millet selection team at the Bambey research station for their contribution to this work. This study is made possible through funding by the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through grants from American People provided to the United States Agency for International Development (USAID) under cooperative agreement number AID-OAA-A-13-00047 and 7200AA-19LE-00005. The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the US Government.

Chapter 4 References

Blümel, M., Dally, N., & Jung, C. (2015). Flowering time regulation in crops—What did we learn from Arabidopsis? *Current Opinion in Biotechnology*, *32*, 121–129. https://doi.org/10.1016/j.copbio.2014.11.023

Bradshaw, W. E., & Holzapfel, C. M. (2017). Natural Variation and Genetics of Photoperiodism in *Wyeomyia smithii*. In S. F. Goodwin (Ed.), *Advances in Genetics* (Vol. 99, pp. 39–71). Academic Press. https://doi.org/10.1016/bs.adgen.2017.09.002

Chen, Z., & Li, Z. (2022). Adaptation and integration of environmental cues to internal flowering network in Arabidopsis thaliana. *Reproduction and Breeding*, *2*(4), 133–137. http://www.sciencedirect.com/science/article/pii/S2667071222000382

Davila-Velderrain, J., Martinez-Garcia, J. C., & Alvarez-Buylla, E. R. (2016). Dynamic network modelling to understand flowering transition and floral patterning. *Journal of Experimental Botany*, *67*(9), 2565–2572. https://doi.org/10.1093/jxb/erw123

Diack, O., Kane, N., Berthouly, C., Guèye, M., Diop, B., Fofana, A., Sy, O., Tall, H., Zekraoui, L., Piquet, M., Couderc, M., Vigouroux, Y., Diouf, D., & Barnaud, A. (2017). New Genetic Insights into Pearl Millet Diversity As Revealed by Characterization of Early-and Late-Flowering Landraces from Senegal. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00818

Diack, O., Kanfany, G., Gueye, M. C., Sy, O., Fofana, A., Tall, H., & Serba, D. D. (2020). *GWAS unveils features between early- and late-flowering pearl millets*. https://doi.org/10.21203/rs.3.rs-25381/v3

Epskamp, S., Costantini, G., Haslbeck, J., Isvoranu, A., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2023). *qgraph: Graph Plotting Methods, Psychometric Data Visualization and Graphical Model Estimation* (Version 1.9.8) [Computer software]. https://cran.r-project.org/web/packages/qgraph/index.html

Faye, A., Barnaud, A., Kane, N. A., Cubry, P., Mariac, C., Burgarella, C., Rhoné, B., Faye, A., Olodo, K. F., Cisse, A., Couderc, M., Dequincey, A., Zekraouï, L., Moussa, D., Tidjani, M., Vigouroux, Y., & Berthouly-Salazar, C. (2022). Genomic footprints of selection in early-and late-flowering pearl millet landraces. *Frontiers in Plant Science*, *13*, 880631. https://doi.org/10.3389/fpls.2022.880631

Ferreira, G. R., Nakaya, H. I., & Costa, L. da F. (2018). Gene regulatory and signalling networks exhibit distinct topological distributions of motifs. *Physical Review E*, *97*(4). https://doi.org/10.1103/PhysRevE.97.042417

Gaudinier, A., & Blackman, B. K. (2020). Evolutionary processes from the perspective of flowering time diversity. *New Phytologist*, *225*(5), 1883–1898. https://doi.org/10.1111/nph.16205

Greenbaum, G., Templeton, A. R., & Bar-David, S. (2016). Inference and Analysis of Population Structure Using Genetic Data and Network Theory. *Genetics*, *202*(4), 1299–1312. https://doi.org/10.1534/genetics.115.182626

Harrell Jr, F. E. (2003). *Hmisc: Harrell Miscellaneous* (p. 5.2-2) [Dataset]. https://doi.org/10.32614/CRAN.package.Hmisc

Itoh, H., Nonoue, Y., Yano, M., & Izawa, T. (2010). A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. *Nature Genetics*, *42*(7), 635–638. https://doi.org/10.1038/ng.606

Kazan, K., & Lyons, R. (2016). The link between flowering time and stress tolerance. *Journal of Experimental Botany*, *67*(1), 47–60. https://doi.org/10.1093/jxb/erv441

Kippes, N., VanGessel, C., Hamilton, J., Akpinar, A., Budak, H., Dubcovsky, J., & Pearce, S. (2020). Effect of phyB and phyC loss-of-function mutations on the wheat transcriptome under short and long day photoperiods. *BMC Plant Biology*, *20*, 297. https://doi.org/10.1186/s12870-020-02506-0

Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S., & Shimamoto, K. (2008). Hd3a and RFT1 are essential for flowering in rice. *Development (Cambridge, England)*, *135*(4), 767–774. https://doi.org/10.1242/dev.008631

Kumar, S., Hash, C. T., Nepolean, T., Satyavathi, C. T., Singh, G., Mahendrakar, M. D., Yadav, R. S., & Srivastava, R. K. (2017). Mapping QTLs Controlling Flowering Time and Important Agronomic Traits in Pearl Millet. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01731

Lakis, G., Navascués, M., Rekima, S., Simon, M., Remigereau, M.-S., Leveugle, M., Takvorian, N., Lamy, F., Depaulis, F., & Robert, T. (2012). Evolution of Neutral and Flowering Genes along Pearl Millet (Pennisetum glaucum) Domestication. *PLOS ONE*, *7*(5), e36642. https://doi.org/10.1371/journal.pone.0036642

Li, J., Li, G., Wang, H., & Wang Deng, X. (2011). Phytochrome Signaling Mechanisms. *The Arabidopsis Book / American Society of Plant Biologists*, *9*, e0148. https://doi.org/10.1199/tab.0148

Li, Y., & Xu, M. (2017). CCT family genes in cereal crops: A current overview. *The Crop Journal*, *5*(6), 449–458. https://doi.org/10.1016/j.cj.2017.07.001

Manning, K., Pelling, R., Higham, T., Schwenniger, J.-L., & Fuller, D. Q. (2011). 4500-Year old domesticated pearl millet (Pennisetum glaucum) from the Tilemsi Valley, Mali: New insights into an alternative cereal domestication pathway. *Journal of Archaeological Science*, *38*(2), 312–322. https://doi.org/10.1016/j.jas.2010.09.007

Mariac, C., Luong, V., Kapran, I., Mamadou, A., Sagnard, F., Deu, M., Chantereau, J., Gerard, B., Ndjeunga, J., Bezançon, G., Pham, J.-L., & Vigouroux, Y. (2006). Diversity of wild and cultivated pearl millet accessions (Pennisetum glaucum [L.] R. Br.) in Niger assessed by microsatellite markers. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *114*(1), 49–58. https://doi.org/10.1007/s00122-006-0409-9

Mockler, T., Yang, H., Yu, X., Parikh, D., Cheng, Y., Dolan, S., & Lin, C. (2003). Regulation of photoperiodic flowering by Arabidopsis photoreceptors. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(4), 2140–2145. https://doi.org/10.1073/pnas.0437826100

Monte, E., Alonso, J. M., Ecker, J. R., Zhang, Y., Li, X., Young, J., Austin-Phillips, S., & Quail, P. H. (2003). Isolation and Characterization of phyC Mutants in Arabidopsis Reveals Complex Crosstalk between Phytochrome Signaling Pathways. *The Plant Cell*, *15*(9), 1962–1980. https://doi.org/10.1105/tpc.012971

Murphy, R. L., Klein, R. R., Morishige, D. T., Brady, J. A., Rooney, W. L., Miller, F. R., Dugas, D. V., Klein, P. E., & Mullet, J. E. (2011). Coincident light and clock regulation of pseudoresponse

regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences*, *108*(39), 16469–16474. https://doi.org/10.1073/pnas.1106212108
Pittendrigh, C. S., & Minis, D. H. (1964). The Entrainment of Circadian Oscillations by Light and Their Role as Photoperiodic Clocks. *The American Naturalist*, *98*(902), 261–294. https://doi.org/10.1086/282327
Saïdou, A.-A., Mariac, C., Luong, V., Pham, J.-L., Bezançon, G., & Vigouroux, Y. (2009). Association Studies Identify Natural Variation at PHYC Linked to Flowering Time and Morphological Variation in Pearl Millet. *Genetics*, *182*(3), 899–910. https://doi.org/10.1534/genetics.109.102756
Song, Y. H., Ito, S., & Imaizumi, T. (2013). Flowering time regulation: Photoperiod- and temperature-sensing in leaves. *Trends in Plant Science*, *18*(10), 575–583. https://doi.org/10.1016/j.tplants.2013.05.003
Su, L., Shan, J.-X., Gao, J.-P., & Lin, H.-X. (2016). OsHAL3, a Blue Light-Responsive Protein, Interacts with the Floral Regulator Hd1 to Activate Flowering in Rice. *Molecular Plant*, *9*(2), 233–244. https://doi.org/10.1016/j.molp.2015.10.009
Sun, C., He, C., Zhong, C., Liu, S., Liu, H., Luo, X., Li, J., Zhang, Y., Guo, Y., Yang, B., Wang, P., & Deng, X. (2022). Bifunctional regulators of photoperiodic flowering in short day plant rice. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.1044790
Tamaki, S., Matsuo, S., Wong, H. L., Yokoi, S., & Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science (New York, N.Y.)*, *316*(5827), 1033–1036. https://doi.org/10.1126/science.1141753
Vicentini, G., Biancucci, M., Mineri, L., Chirivì, D., Giaume, F., Miao, Y., Kyozuka, J., Brambilla, V., Betti, C., & Fornara, F. (2023). Environmental control of rice flowering time. *Plant Communications*, *4*(5), 100610. https://doi.org/10.1016/j.xplc.2023.100610
Vigouroux, Y., Mariac, C., Mita, S. D., Pham, J.-L., Gérard, B., Kapran, I., Sagnard, F., Deu, M., Chantereau, J., Ali, A., Ndjeunga, J., Luong, V., Thuillet, A.-C., Saïdou, A.-A., & Bezançon, G. (2011). Selection for Earlier Flowering Crop Associated with Climatic Variations in the Sahel. *PLOS ONE*, *6*(5), e19563. https://doi.org/10.1371/journal.pone.0019563
Wang, H., Jia, G., Zhang, N., Zhi, H., Xing, L., Zhang, H., Sui, Y., Tang, S., Li, M., Zhang, H., Feng, B., Wu, C., & Diao, X. (2022). Domestication-associated PHYTOCHROME C is a flowering time repressor and a key factor determining Setaria as a short-day plant. *New Phytologist*, *236*(5), 1809–1823. https://doi.org/10.1111/nph.18493
Wang, X., Zhou, P., Huang, R., Zhang, J., & Ouyang, X. (2021). A Daylength Recognition Model of Photoperiodic Flowering. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.778515
Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., Freidank, M., Cai, J., & Protivinsky, T. (2024). *corrplot: Visualization of a Correlation Matrix* (Version 0.95) [Computer software]. https://cran.r-project.org/web/packages/corrplot/index.html
Wolabu, T. W., Zhang, F., Niu, L., Kalve, S., Bhatnagar-Mathur, P., Muszynski, M. G., & Tadege, M. (2016). Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. *New Phytologist*, *210*(3), 946–959. https://doi.org/10.1111/nph.13834

Xie, X., Kagawa, T., & Takano, M. (2014). The Phytochrome B/Phytochrome C Heterodimer Is Necessary for Phytochrome C-Mediated Responses in Rice Seedlings. *PLoS ONE*, *9*(5), e97264. https://doi.org/10.1371/journal.pone.0097264

Yang, S., Weers, B. D., Morishige, D. T., & Mullet, J. E. (2014). CONSTANS is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biology*, *14*(1), 148. https://doi.org/10.1186/1471-2229-14-148

Yang, X.-F., Li, X.-M., Ingvarsson, P. K., Xi, C., & Liao, W.-J. (2024). Molecular mechanisms of flowering time differentiation revealed by transcriptomic sequencing and de novo analysis in Chinese invasive populations of Ambrosia artemisiifolia. *BMC Plant Biology*, *24*(1), 1106. https://doi.org/10.1186/s12870-024-05830-x

Zheng, T., Sun, J., Zhou, S., Chen, S., Lu, J., Cui, S., Tian, Y., Zhang, H., Cai, M., Zhu, S., Wu, M., Wang, Y., Jiang, L., Zhai, H., Wang, H., & Wan, J. (2019). Post-transcriptional regulation of Ghd7 protein stability by phytochrome and OsGI in photoperiodic control of flowering in rice. *New Phytologist*, *224*(1), 306–320. https://doi.org/10.1111/nph.16010

Chapter 4 Figures and Tables

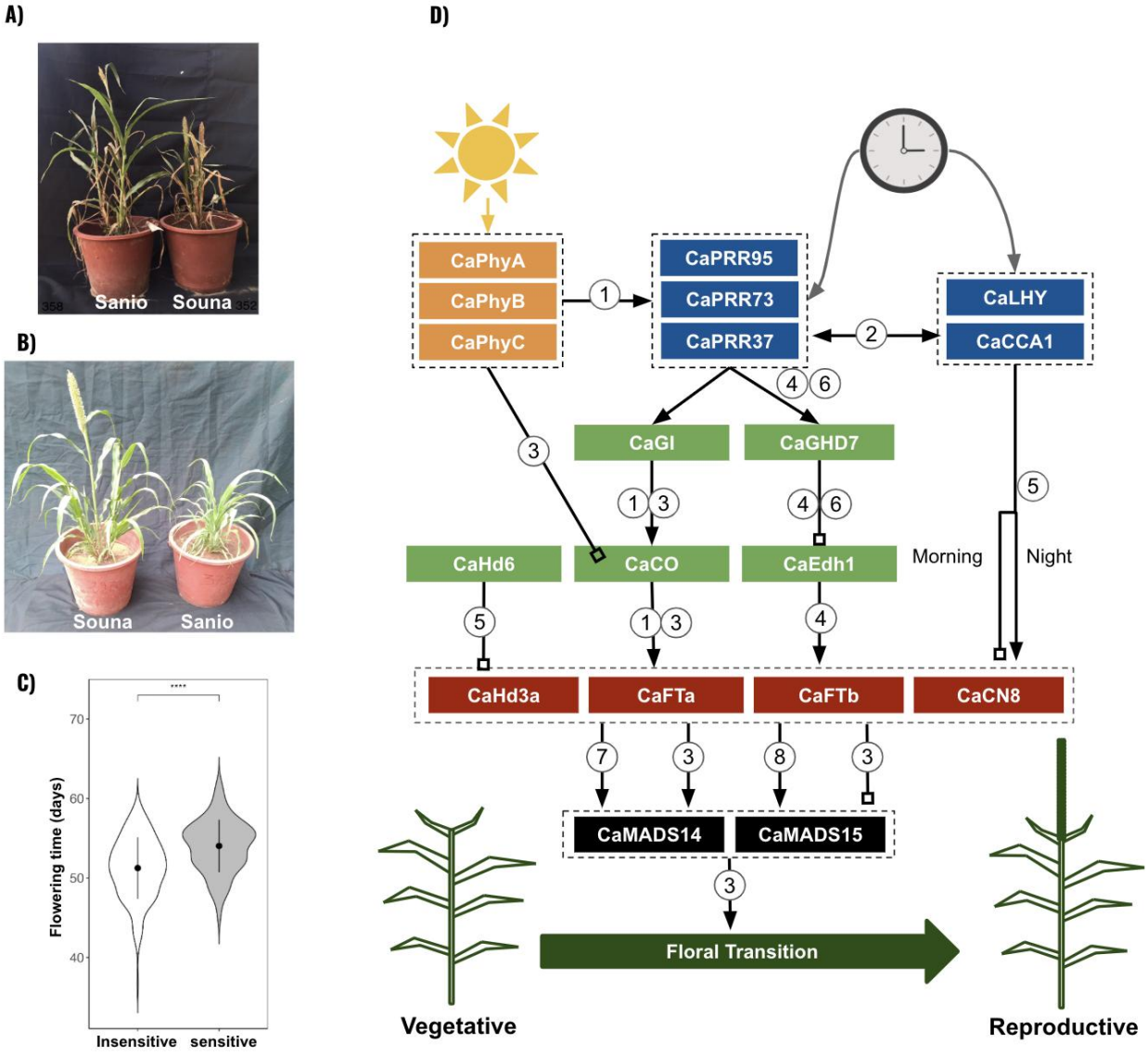


Figure 1 - Hypothetical model of conservative regulation of floral transition in pearl millet. Image of Souna vs Sanio phenotypes at flowering (A) and maturity stages (B). Phenotypic distribution for photoperiod insensitive *vs* sensitive lines in the PMIGAP diversity panel (C). Hypothetical model of flowering period control based on previous observations (D). Genes from the same family are framed in a box with dotted lines. Arrows highlight positive regulatory relationships and bars indicate negative regulatory relationships. Pearl millet flowering is triggered by environmental cues (light and a circadian clock). Light regulates phytochromes (CaPHA, CaPhyB and CaPhyC) and the circadian clock regulates pseudo-response regulatory proteins (CaPRR37, CaPRR73, CaPRR97), CaLHY and CaCCA1. GI activates CO which

induces FT genes (caHd3a, CaCN8 and CaFTa) which in turn induces CaMADS14. The independent GI pathways conclude the two activators CaHd6 and CaEhd1 and the inhibitors CaFTb. Literature references are as follows: (1) Dong and al., 2012 ; (2) Song and al., 2010; (3) Wang and al., 2022; (4) Casto and al., 2019; (5) Fujiwara and al., 2008; (6) Pautler and al., 2013; (7) Sohail and al., 2023; (8) Meng and al., 2011.

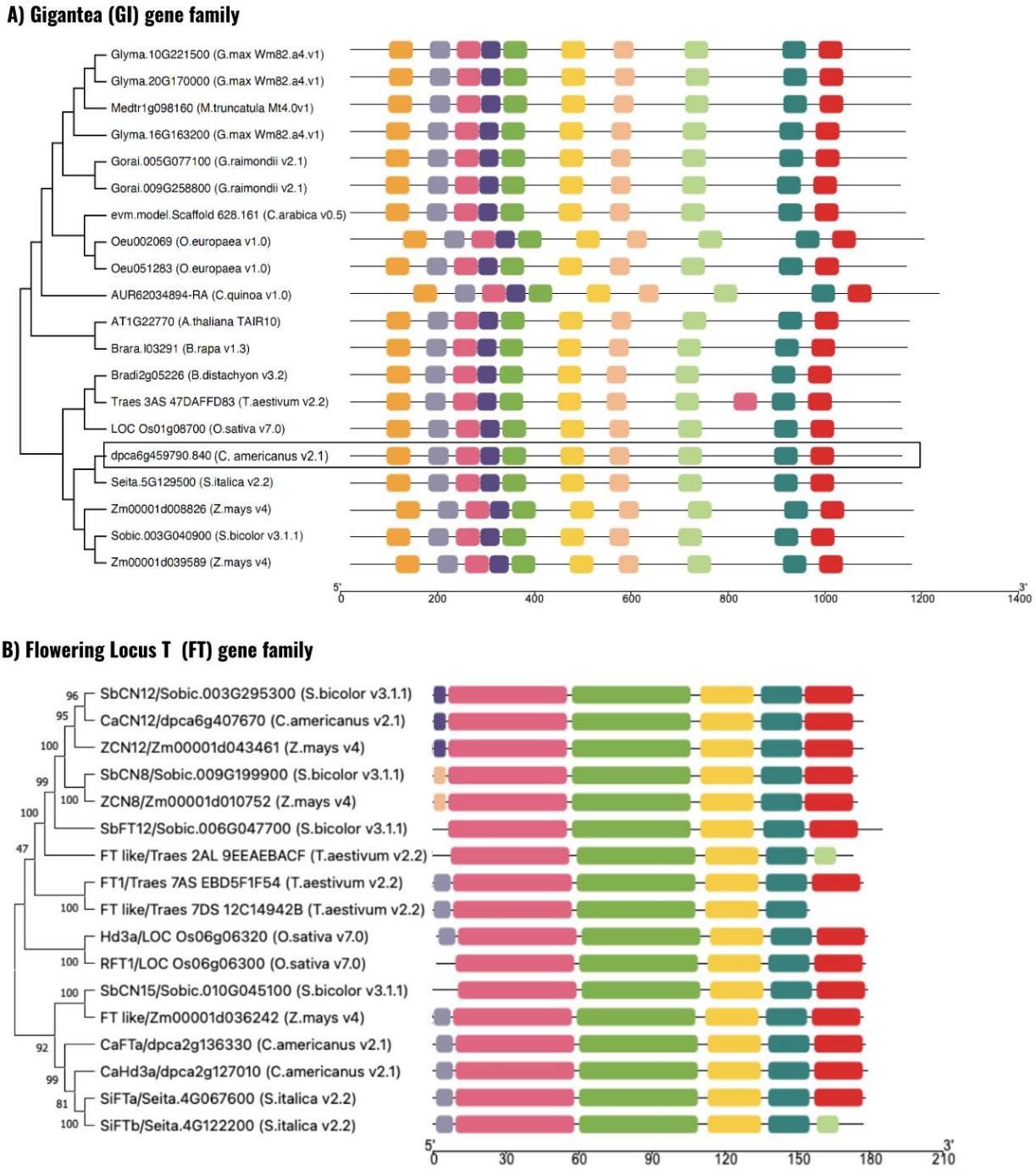


Figure 2. Conserved protein domain in candidate flowering genes found in pearl millet genome. Phylogenetic analysis of CaGI (A) and CaFTs (B) found in reference genome v2.1 (Ramu and al. 2023) and their related proteins from various cereal crops and, respectively. Phylogenetic analysis with MEGA11 and protein motif analysis using MEME and TBtools.

Phytozome and ncbi accession numbers are indicated. The differently colored boxes indicated the different protein motifs found in the candidate genes compared to their orthologs.

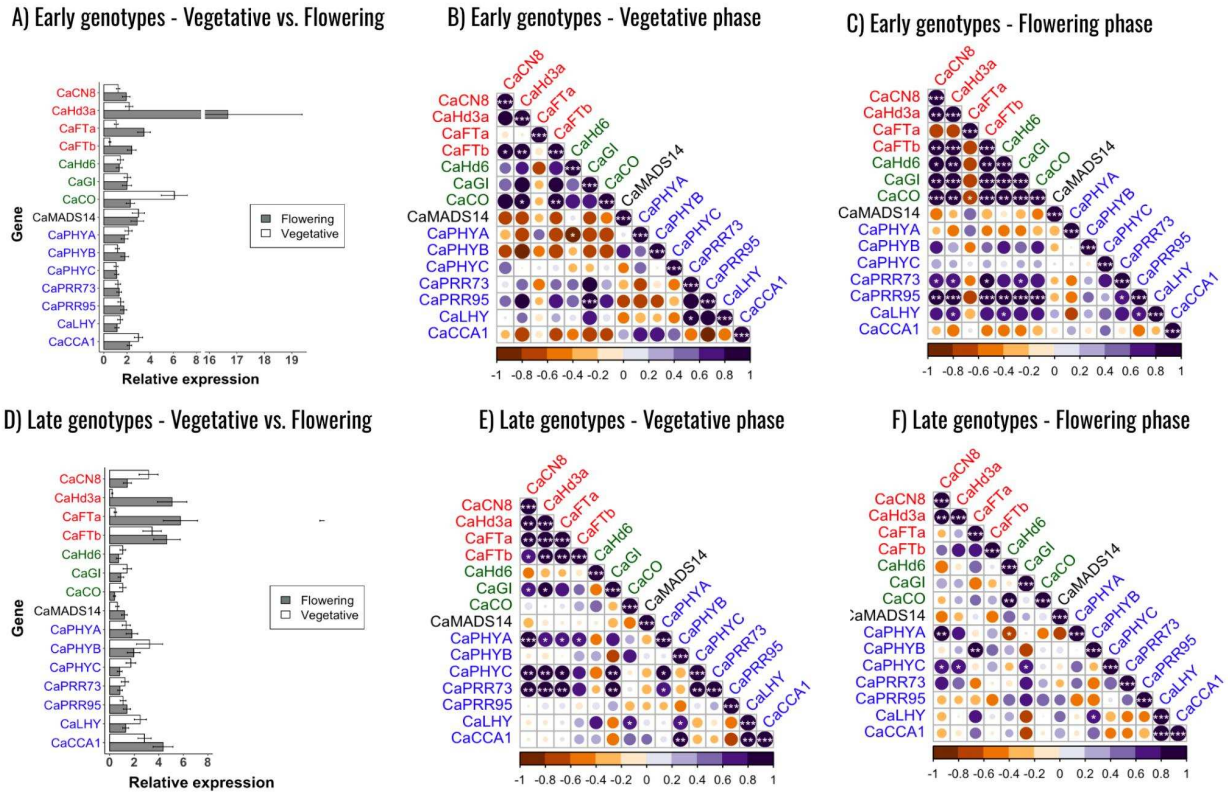


Figure 3 - Photoreceptor signals trigger early floral transition via CaCO and CaHd3a, while CaFTa and CaFTb act later in development. (A) Expression profiles of flowering-time genes categorized by function: Florigens (red), Photoperiod/clock components (blue), Signal integrators (green), and Floral development genes (black), measured by qPCR in early- and late-flowering genotypes at vegetative and flowering stages. All expression values are shown as relative expression levels normalized to the reference gene CaTIP41, and represent the mean of three biological replicates. Expression values were calculated using the 2^−ΔCt method, where ΔCt = Ct(target) − Ct(reference), and are thus relative, not absolute expression values.(B–C) Gene co-expression networks inferred from correlation matrices for early-flowering genotypes at the vegetative (B) and flowering (C) stages. (E–F) Same analysis for late-flowering genotypes at vegetative (E) and flowering (F) stages. Significance of correlations is indicated by asterisks: *p < 0.05; **p < 0.01; ***p < 0.001.

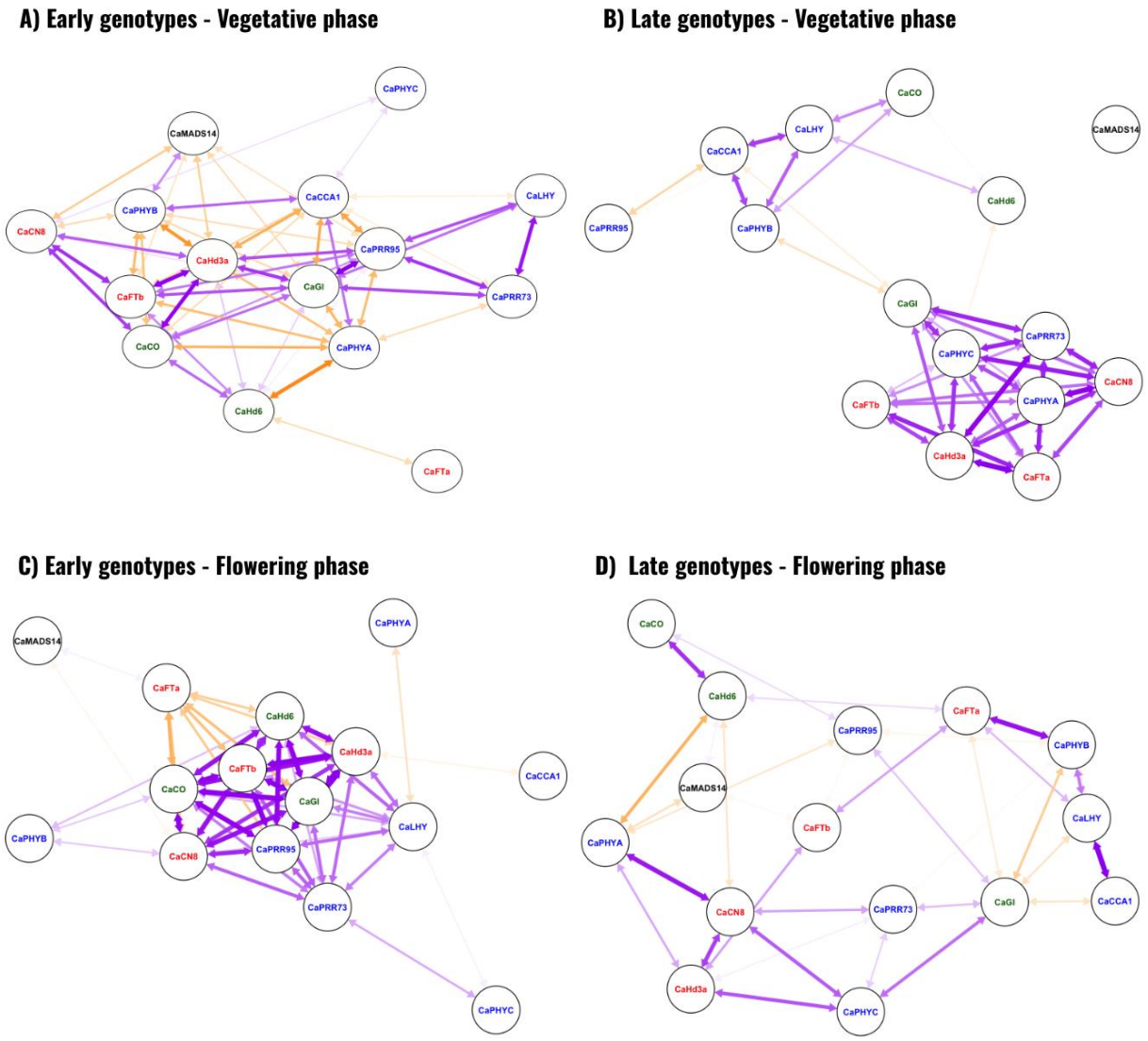


Figure 4 - Co-expression networks show stronger coordination between photoperiod signals and florigen genes in early-flowering (Souna) genotypes than in late-flowering (Sanio) genotypes. Networks were generated using the *qgraph* R package from Pearson correlation matrices of gene expression during the vegetative (A, B) and flowering (C, D) stages in early (A, C) and late (B, D) genotypes. Nodes represent genes, colored by function: Florigen (red), Photoperiod/Clock (blue), Integrator (green), Floral Development (black). Edges represent gene–gene correlations: purple for positive, orange for negative. Darker, thicker edges indicate stronger correlations. Edge length is layout-dependent but reflects interaction strength, shorter edges typically link more tightly co-expressed genes. The more compact network in (C) suggests stronger regulatory coordination compared to the fragmented layout in (D).

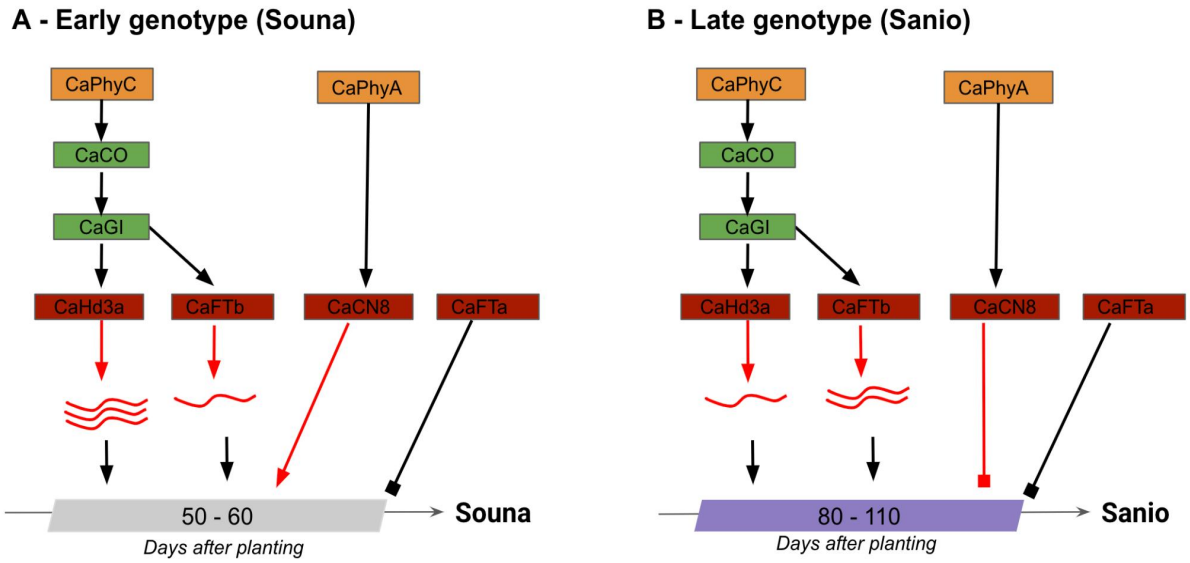


Figure 5 - Proposed models of flowering regulation of the contrasting response of Souna (A) and Sanio (B) genotypes to photoperiodic signals. Gene expression levels are represented using arc symbols below each gene. Boxes represent genes, colored by function: Florigen (red), Photoperiod/Clock (blue), Integrator (green), Floral Development (black). The differences between Souna and Sanio are highlighted by red color lines.

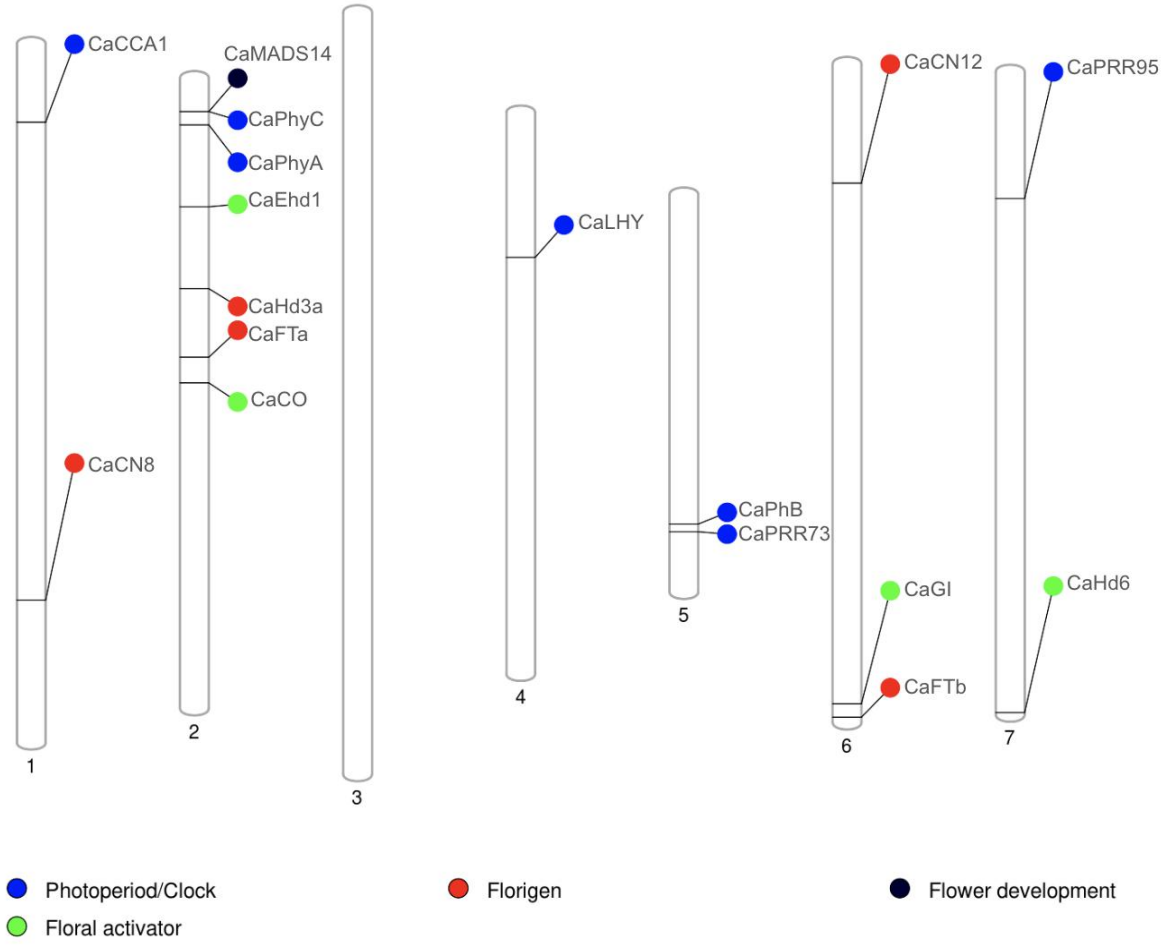


Figure S1. The core flowering genes were conserved in the pearl millet genome during its speciation. The physical location of candidate orthologs of core flowering time genes is represented on all seven chromosomes. Color coding was used to indicate the putative functional role of candidate genes in flowering time regulation.

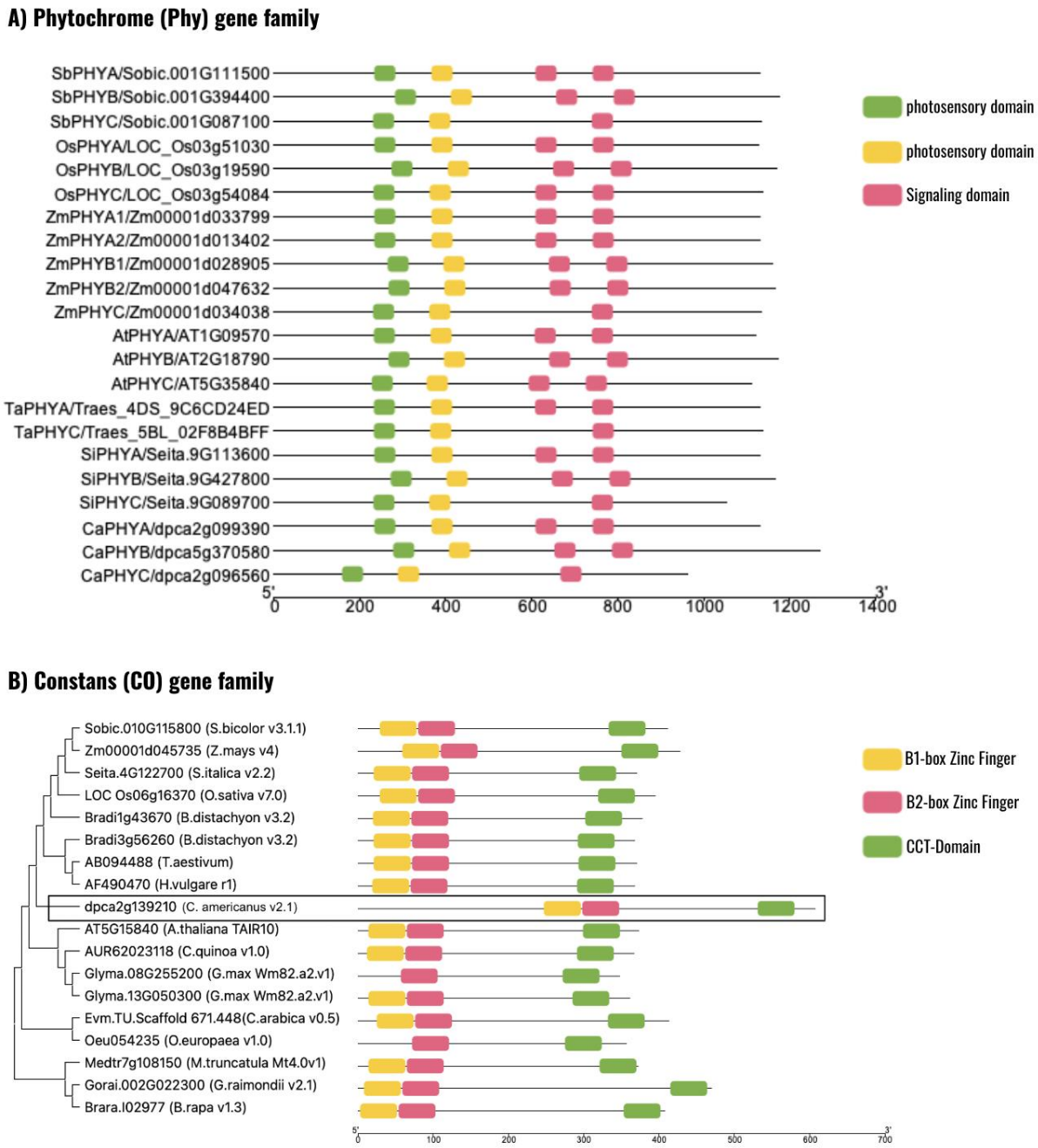


Figure S2. Conserved protein domain in CONSTANS (CO) and Phytochrome (Phy) gene families found in pearl millet genome. Phylogenetic analysis of CaCO and CaPhys found in reference genome v2.2 (Ramu and al. 2023) and its related proteins from various cereal crops (A) and (B), respectively. Phylogenetic analysis with MEGA11 and protein motif analysis of PglCO using MEME and TBtools. Phytozome and ncbi accession numbers are indicated. For the

CO gene family the yellow and red boxes indicate conserved domains corresponding to the B1 and B2-box zinc fingers required for protein binding. Green boxes represent conserved domains corresponding to the CCT domain required for zinc ion binding.

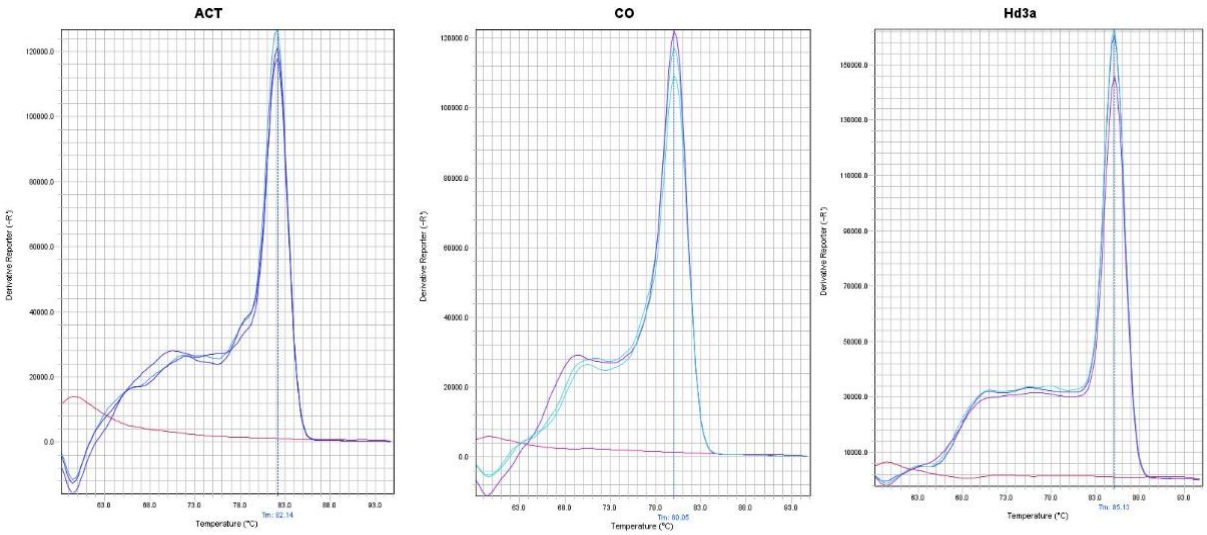


Figure S3. Image of specific primers designed and used in the qPCR essay in this study.

Table S1. Accession numbers of core flowering time gene orthologs in cereal crops used in this study

Gene symbol	Foxtail millet	Maize	Sorghum	Rice	Wheat
GI	Seita.5G129500	Zm00001d008826	Sobic.003G040900	LOC_Os01g08700	Traes_3B_A90EC49F7 Traes_3B_5F7FFAC8C
CO	Seita.4G122700	Zm00001d045735	Sobic.010G115800		-
LHY	Seita.6G055700	Zm00001d024546	Sobic.007G047400	LOC_Os08g06110	-
CN12	Seita.4G122700	Zm00001d043461	Sobic.003G295300	LOC_Os06g16370	AB094488
FTa	Seita.4G067600	-	-	-	-
CCA1	-		Sobic.007G047400	-	-
PhyB	Seita.9G427800	Zm00001d028905 Zm00001d047632	Sobic.001G394400	LOC_Os03g19590	-
PHYC	Seita.9G089700	Zm00001d034038	Sobic.002G313800	LOC_Os03g54084	Traes_5BL_02F8B4BFF
PhyA	Seita.9G113600	Zm00001d033799 Zm00001d008542 Zm00001d013402	Sobic.001G111500	LOC_Os03g51030	Traes_4DS_9C6CD24ED
CN8	-		Sobic.009G199900	-	-
FTb	Seita.4G122200		-	-	-
Hd3a	-	-	-	LOC_Os06g06320	-
RFT1	-	-	-	LOC_Os06g06300	-
Hd6	-	-	-		-
CN15	Seita.4G067600	Zm00001d036242	Sobic.010G045100	LOC_Os06g06300	Traes_2AL_9EEAEBACF
Ehd1	Seita.1G351300	Zm00001d018380	Sobic.001G227900	LOC_Os10g32600	-
Ghd7		Zm00001d012441	Sobic.006G004400	LOC_Os07g15770	-

Dashes (–) indicate that no orthologous gene was identified in the species, either because the gene has not yet been cloned or could not be found in available NCBI or Phytozome databases.

Table S2: Physical location of candidate orthologs of core flowering time genes found in the pearl millet reference genome v2.1 (Ramu and al. 2023)

Gene	Chromosome	Start	End	Phenotype	Tift 2023-gene
PHYC	Chr02	14340819	14338788	Photoperiod	dpca2g096560
PHYB	Chr05	140946099	140944069	Photoperiod	dpca5g370580
PHYA	Chr02	20092764	20094826	Photoperiod	dpca2g099390
PRR73	Chr05	144021910	144022676	Photoperiod	dpca5g372370
PRR95	Chr07	54324457	54325200	Photoperiod	dpca7g487250
CCA1	Chr01	33430150	33430255	Clock	dpca1g012730
GI	Chr06	273574553	273573049	Photoperiod/Clock	dpca6g459790
CO	Chr02	130142482	130142985	Photoperiod/Clock	dpca2g139210
LHY	Chr04	61984781	61986161	Clock	dpca4g284610
Ehd1	Chr02	54912085	54912435	Photoperiod	dpca2g114190
Hd3a	Chr02	89842073	89842299	Florigen	dpca2g127010
Hd6	Chr07	273859554	273859717	Flower induction	dpca7g541340
LHSI	Chr05	154337635	154337819	Inflorescence meristem	Unannotated
MADS14	Chr02	14160297	14160481	Flower transition	dpca2g096490
MADS15	Chr07	273920948	273920764	Flower transition	Unannotated
MADS11	Chr02	108430991	108431184	Flower transition	dpca2g133390
PgCN8	Chr01	237844093	237843862	Florigen	Unannotated
PgCN12	Chr06	50767493	50767263	Florigen	dpca6g407670
PgFTa	Chr02	119372277	119372047	Florigen	dpca2g136330
PgFTb	Chr06	279365265	279365474	Florigen	Unannotated

Table S5. The primer is used for expression of pearl millet putative flowering genes using QPCR analysis

Gene name	Forward sequence	Reverse sequence	Product size (bp)
CaGI	AGCTGTTCCTGCACTTCTG	ACCGGCTCTCATTGACATTG	98
CaCO	TCGGCTACAATTCATACTGCG	AGCTCCTGTTCTTGCCTTC	88
CaLHY	AGAAGTCACGCCTCAGAAGTTC	AATTGTACATTGCCATTACTCAGTG	178
CaCN12	GTGTGATACGAGATGTGTTGGA	TGCCCCCAACATCAACTC	153
CaFTa	CCAGGACCGTAGCCAAC	CTGGCAGTTGAAGTAGACGG	89
CaCCA1	ATCACCAGTTCAGAGTCAACAA	GATGAGGATGGACCAGGAAAG	96
CaPhyB	CCCACTGAGTCTCAAATTACGG	TCCCGAATAAGCTGCATATCC	120
CaPHYC	TTCGTTCCTTTCCCGCTAAG	GCCTCTGTTCCTCTGTCAAC	89
CaPhyA	GCAGTAGATGGCAATGGATTG	GAGGAATGAGTGGGTTTGGAT	98
CaEhd1	CTTAAAGTTTGACATCCCTACCTG	GCTTCAGTGATCCCCATGG	105
CaLHSI	ATCGAGAACATATCAGCCGG	AGAAGATGATGAGCGCGAC	92
CaCN8	GGCATCCACAGGATGGTATTTA	TGAAGTTTGTGGCGCATTTC	83
CaFTb	TTCAGCACCCGTGACTTC	ATTCTCCTGCCACCAGTTC	101
CaPRR73	TCAGAATCGGTTCCCTCAAC	TGGTCCCCTTGATCTTCTTG	174
CaPRR95	GAATCAACCCCACAAGCAAC	ATGCCGCACCTCATCTTTAG	165
CaHd3a	CTCCAAGCCCAAGTGACC	ACAGCACTAACACGAATCGGT	153
CaHd6	TATGCACCGAGATGTGAAGC	TGCAAGTCCACAAGAAGCTC	165
ACT	TGCTCAGTGGAGGATCTACTAT	CTGGTGGTGCAATCACTTTAAC	108
UBQ5	GTACACCAAGCCCAAGAAGA	GTCGTCGACCTTGTAGAACTG	79
PP2A	TGAGAGCAGACAAATCACTCAA	AAGAGCTGTGAGAGGCAAATAA	121
TIP41	GGTTCTGAACTCAGGCACTAC	GAAAGGGCAACAAGGTCAATC	96

Table S6. Information on the origin, pedigree and flowering time of the genotypes used in this study

DESIGNATION	Pedigree	Country	Sample.ID	Biological.status	Flowering (days)	Ecotype
ICML 197281	IP-4927	Senegal	PMiGAP257	Breeding line	42	early
ICML 197352	IP-9347	Ghana	PMiGAP056	Breeding line	46	early
ICML 197347	IP-8949	Togo	PMiGAP090	Breeding line	56	early
ICML 197297	IP-5923	Senegal	PMiGAP226	Breeding line	63	late
ICML 197373	IP-10701	Mali	PMiGAP215	Landrace	70	late
ICML 197458	IP-21020	Nigeria	PMiGAP173	Landrace	67	late

Chapter 5: General conclusion

This thesis integrates population genetics, molecular biology, and quantitative trait dissection to develop a strategic framework for accelerating genetic gain in millet breeding. The core question addressed is how to improve target traits such as grain yield while conserving elite backgrounds for adaptive traits like flowering time, especially in resource-limited breeding programs.

In Chapter 2, we established that Quantitative Genomic Identity (QGI) provides the most accurate method for inferring elite relationships based on known QTL, offering a foundation for designing optimal crosses. This insight highlighted the importance of identifying the genetic architecture of key traits before defining elite relationships. Thus, Chapters 3 and 4 focused on uncovering the genetic and regulatory basis of flowering time, a pivotal trait for drought adaptation in pearl millet.

In Chapter 3, we used a forward genetics approach through GWAS in West African germplasm to identify an oligogenic architecture for flowering time, with major effects from loci including *Phytochrome C*. These results revealed that while local ecotypes show clear phenotypic divergence, they often share genetic backgrounds at regional scales, providing opportunities for cross-border breeding synergy.

Building on this, Chapter 4 used a reverse genetics approach, combining phylogenetic analysis and gene expression profiling in contrasting ecotypes. This revealed a GI-dependent regulatory pathway, modulated by photoreceptors like *CaPhyC*, that governs the timing of floral transition.

The findings culminated in a first regulatory model for flowering in pearl millet, highlighting ecotype-specific expression dynamics and compensation mechanisms.

Collectively, these studies deliver a full-circle strategy from trait architecture discovery to elite gene pool definition for guiding trait introgression in millet. The approach enables breeders to maintain elite performance for adaptive traits while driving genetic gain for productivity traits like yield. This thesis emphasizes the power of integrating molecular inference, trait dissection, and regulatory network analysis to build climate-resilient, high-performing varieties tailored to the unique challenges of the Sahel and beyond.