UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

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

Genetic determinism of seed production in alfalfa used as living mulch

Master's Thesis

Erasmus Mundus Master Programme in Plant Breeding - emPLANT +

AUTHOR: Guacaneme Moreno, Martín Andrés

Tutor: Prohens Tomás, Jaime

Experimental director: Julier, Bernadette

ACADEMIC YEAR: 2023/2024

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

INRAE

Universitat Politècnica de València

Institute for the Conservation and Improvement of Valencian Agrobiodiversity (COMAV)

Erasmus Mundus Master's degree in Plant Breeding (emPLANT+)

Genetic Determinism of Seed Production in Alfalfa used as Living Mulch

Academic Year: September 2023 - June 2024

Martin Guacaneme

Supervisor: Bernadette Julier (INRAE)
Co-supervisor: Zineb El Ghazzal (INRAE)
Co-supervisor: Marie Pégard (INRAE)
Academic Supervisor: Jaime Prohens (UPV)

Location and Date of Submission: València, July 2024

Genetic Determinism of Seed Production in Alfalfa used as Living Mulch

Resumen

Este estudio investiga los determinantes genéticos de la producción de semillas en la alfalfa (*Medicago sativa*) utilizada como mantillo vivo, en el contexto de la transición agroecológica. Utilizando un panel diverso compuesto por una población F1 y 30 poblaciones de alfalfa, incluyendo tipos cultivados y silvestres, analizamos un total de 1,600 plantas. Cada planta fue fenotipada para siete rasgos de producción de semillas y genotipada utilizando la Secuenciación por Genotipado (GBS) para obtener frecuencias alélicas de SNP. Utilizando los datos de GBS, la estructura poblacional del panel fue analizada mediante Análisis de Componentes Principales (PCA) y evaluación de Desequilibrio de Ligamiento (LD). Las varianzas genéticas se determinaron usando el modelo REML y se utilizaron para calcular la heredabilidad en sentido amplio. Los Estudios de Asociación del Genoma Completo (GWAS) a través de un Modelo Mixto Multilocus (MLMM) se utilizaron como metodología para detectar QTLs candidatos. Se observó una estructura genética en el panel, con las poblaciones diploides distanciándose de las tetraploides. Por otro lado, se encontró que F1 presentaba una alta estructura genética. Por lo tanto, tanto las poblaciones tetraploides F1 como no F1 fueron tratadas por separado en los Estudios de Asociación. Se identificaron dieciséis marcadores genéticos asociados con rasgos de producción de semillas; la mayoría de estos marcadores estaban vinculados a genes anotados. Estos resultados, junto con la identificación de QTLs candidatos significativos, subrayan variaciones genéticas sustanciales que podrían informar estrategias de mejoramiento para desarrollar variedades de alfalfa optimizadas para su uso como mantillo vivo con alto potencial de semillas.

Abstract

This study investigates the genetic determinants of seed production in alfalfa (*Medicago sativa*) utilized as living mulch, in the context of agroecological transition. Utilizing a diverse panel comprising one F1 and 30 alfalfa populations, including both cultivated and wild types, we analyzed a total of 1,600 plants. Each plant was phenotyped for seven seed production traits and genotyped using Genotyping-by-Sequencing (GBS) to obtain SNP allele frequencies. Using the GBS data, the panel's population structure was analyzed through Principal Component Analysis (PCA) and Linkage Disequilibrium (LD) assessment. Genetic variances were determined using the REML model and were used to calculate broad-sense heritability. Genome-Wide Association Studies (GWAS) through a Multi-Locus Mixed Model (MLMM) were used as the methodology to detect candidate QTLs. A genetic structure was observed in the panel, with diploid populations distancing from tetraploid populations. On the other hand, F1 was found to present high genetic structure. Subsequently, both F1 and non F1 tetraploid populations were treated separately in the Association Studies. Sixteen genetic markers associated with seed production traits were identified; most of these markers were linked to annotated genes. These results, along with the identification of significant candidate QTLs, underscore substantial genetic variations that could inform breeding strategies for developing optimized alfalfa varieties for use as living mulch with high seed potential.

Palabras Clave: lucerna, GWAS, producción de semilla, QTL, asociación de cultivos
Keywords: alfalfa, GWAS, seed production, QTL, living mulch, companion plant

Author:	Martin Guaceneme
Location and date:	Valencia, July 2024
External Supervisor:	Bernadette Julier
Academic supervisor:	Jaime Prohens

Acknowledgements

I am profoundly grateful to Bernadette Julier and Zineb El Ghazzal, PhD student at INRAE, for their invaluable guidance and expertise throughout the development of this work. Their detailed insights and constructive feedback were instrumental in achieving the success of this project. I also wish to thank Marie Pégard, Core Scientist at INRAE, for her contributions with the R routines used throughout this TFM. My thanks extend to INRAE, especially the Grasslands and Fodder Plants Research Unit at Lusignan, for providing the necessary support and resources crucial for my research. I also thank Irving Arcía, a PhD student working on the alfalfa pangenome, from whom I borrowed the figures used in the section on the history of alfalfa. A special note of appreciation goes to Bernadette Julier and Zineb El Ghazzal for their thorough review and suggestions that greatly enhanced the quality of this TFM.

I extend my gratitude towards Marco de la Vega for his unconditional support during these months of internship. Finally I want to acknowledge El Presidente oak tree, king of the Saint-Sauvant silvicultural forest, for allowing us to visit his domain several times during my stay in the region.

Preface

The table below shows the degree of relation of this master's thesis work (TFM, acronym in Spanish for Trabajo de Fin de Máster) with the United Nations Sustainable Development Goals (SDGs) of the 2030 Agenda.

Sustainable Development Goal	High	Medium	Low	NA
1. End poverty.				x
2. Zero hunger.			x	
3. Good health and well-being.				x
4. Quality education.				x
5. Gender equality.				x
6. Clean water and sanitation.		x		
7. Affordable and clean energy.				x
8. Decent work and economic growth.				x
9. Industry, innovation, and infrastructure.			x	
10. Reduced inequalities.				x
11. Sustainable cities and communities.			x	
12. Responsible consumption and production.			x	
13. Climate action.			x	
14. Life below water.				x
15. Life on land.	x			
16. Peace, justice, and strong institutions.				x
17. Partnerships for the goals.				x

The goal most directly addressed by this work is Goal 15, "Life on Land." This goal is dedicated to protecting, restoring, and promoting the sustainable use of terrestrial ecosystems, managing forests sustainably, combating desertification, halting and reversing land degradation, and halting biodiversity loss.

This TFM is an integral component of the MoBiDiv project, which is geared towards developing genetically diverse alfalfa varieties that are optimized for intra-plot diversification. This directly supports the transition to pesticide-free agriculture, enhancing the sustainability of agricultural practices. When employed as a living mulch, alfalfa effectively suppresses weeds, enriches the soil with organic nitrogen —eliminating the need for synthetic fertilizers— and enhances soil health and structure. Furthermore, the presence of alfalfa varieties coupled with cash crops increases the biodiversity within agricultural systems, which is a direct action towards achieving SDG 15's aim of promoting sustainable ecosystem use and halting biodiversity loss.

Understanding the genetic determinants of seed production in alfalfa is crucial because it helps in selecting varieties that are more efficient in their roles as living mulch. Intercropping practices contribute to maintaining ecosystem balance and also enhances agricultural sustainability by reducing dependency on chemical inputs and improving soil and plant health. Thus, my TFM aligns with the objectives of SDG 15 by contributing to the restoration and sustainable management of terrestrial ecosystems through innovative agricultural practices.

Contents

1 Introduction

1.1 Context

Intensive agricultural practices, including short crop rotations, mechanization, monoculture, and the excessive use of fertilizers and pesticides, have led to severe environmental challenges. These include soil degradation, water contamination, loss of biodiversity, increased disease incidence, and a decline in long-term agricultural sustainability (Marcos et al., 2020). Monoculture, a particularly harmful practice, not only increases soil erosion and pollutes water resources but also elevates atmospheric carbon levels and diminishes biodiversity (Demirdogen et al., 2023). It affects a wide range of agricultural ecosystems, from artificially created 'cultivated forests' (Karancsi, 2010) to cash crops such as wheat and cotton (Demirdogen et al., 2023). To mitigate these issues, sustainable practices such as crop rotation and multicropping have been advocated as effective alternatives.

Crop rotation, wherein farmers alternate crop types on the same land from year to year or within the same year, not only boosts crop yields but also stabilizes them, reduces pest and disease outbreaks, and ultimately enhances farmers' financial results (Demirdogen et al., 2023). Multicropping, which involves growing multiple crops in the same space, utilizes methods such as sequential cropping (growing crops one after another) and intercropping (growing crops simultaneously in close proximity). These practices aim to optimize the use of resources — such as light, water, and nutrients— and increase yield while minimizing risks associated with pests and diseases (Deb, 2021).

Forage intercropping presents a viable agroecological practice. The use of legume associations, exemplified by systems such as wheat-faba bean, wheat-pea, barley-pea, and maize-cowpea, leverages biological nitrogen fixation from legumes, reducing reliance on synthetic fertilizers (Mu et al., 2023).

A specialized form of intercropping, companion planting, strategically pairs specific plants to benefit one to the other in terms of pest control, pollination, habitat provision for beneficial organisms, space maximization, and overall plant health and soil quality. For instance, some plants naturally repel pests, reducing the need for chemical interventions (Huss et al., 2022).

An interesting variation of companion planting involves the use of living mulches (or living cover crops). Living mulches are a type of cover crop sown either before or simultaneously with the cash crop and kept alive as ground cover for the entire growing period. When the living mulch consists of perennial plants, it can potentially be preserved over several years without requiring reseeding (Cougnon et al., 2022).

In France, the usage of living mulches is gaining traction (Carof, 2006; Labreuche et al., 2017). Alfalfa, in particular, has been reported as an effective living mulch to combat weeds, enrich the soil with nitrogen, and enhance soil structure (Petit & Aubry, 2012). These advantages enable more environmentally sustainable agronomic practices.

1.2 Alfalfa

Medicago sativa L. (alfalfa) stands as an important legume perennial species, widely cultivated for forage (Annicchiarico et al., 2015). It is recognized as the highest protein producer under temperate climates (Pégard et al., 2023), contributing significantly to protein auton-

omy through its efficient atmospheric nitrogen fixation. Remarkably, alfalfa achieves unrivaled protein productivity per hectare in temperate zones, reaching up to four tons of protein per hectare. It also boasts an adequate energy value and enhances the structure and composition of soils (Mei et al., 2022; Osterholz et al., 2019).

Alfalfa's robust deep root system renders it drought-tolerant, capable of accessing water from deep within the soil, thus ensuring its resilience (Jefferson & Cutforth, 2005). This deep rooting also facilitates nitrogen fixation and enhances water quality. Additionally, it interrupts the life cycles of weeds, pests, and pathogens in annual crops, further contributing to improved carbon storage in the soil (Fernandez et al., 2019; Meiss et al., 2010).

1.2.1 History and Origin

The history and origin of alfalfa are intricately linked to human civilization, reflecting a journey that spans continents and millennia. Initially believed to have been cultivated as early as 9,000 years ago in the Near East to Central Asia, alfalfa's domestication is rooted in regions rich in agricultural innovation (Hanson et al., 1988). Various scholars, including Prosperi et al. (2014), have traced the earliest written records of alfalfa to its introduction to Greece by the Medes, followed by its spread across Italy and the wider Roman Empire (Figure 1).

The Middle Ages saw a decline in the cultivation of alfalfa in Europe, but it resurged when reintroduced in Spain by the Moors, eventually spreading across Europe and later to the Americas by the Spanish in the sixteenth century. After the Spanish arrived in South America, alfalfa was introduced to the new continent in the 18th century. This expansion continued into North America through California in the early nineteenth century, marked as "Chilean clover" (Prosperi et al., 2014). The introduction of alfalfa to North America and Oceania occurred considerably later, and the sources and routes taken for this introduction are well documented (Şakiroğlu & İlhan, 2021)

The debate over alfalfa's precise origins remains open, with multiple potential centers of origin proposed, including Eastern Anatolia, Iran, Armenia, Afghanistan, and Central Asia (Lesins & Lesins, 1979). Each of these regions presents unique contributions to alfalfa's genetic diversity, shaped by natural and artificial selection to adapt the crop to varied ecological conditions. Central Asia, in particular, is noted for its role as a secondary center of diversity, contributing genetic resistance to diseases like bacterial wilt and pests such as the blue alfalfa aphid.

Prosperi et al. (2014) emphasize that despite a significant domestication bottleneck, which reduced genetic variability, alfalfa maintains considerable genetic diversity, particularly in wild populations found in its areas of origin. These wild forms, particularly adapted to their local environments, offer valuable genetic traits such as drought tolerance and prostrate growth habits, which are crucial for alfalfa's adaptation to new and changing environments.

1.2.2 Taxonomy

Alfalfa belongs to the major genus *Medicago* within the Fabaceae family. It is part of a taxonomic continuum that encompasses both the cultivated crop alfalfa varieties and its closely related wild taxa, all of which can interbreed and produce fertile offspring. This taxonomic continuum is referred to as the *Medicago sativa* species complex (Şakiroğlu & İlhan, 2021).

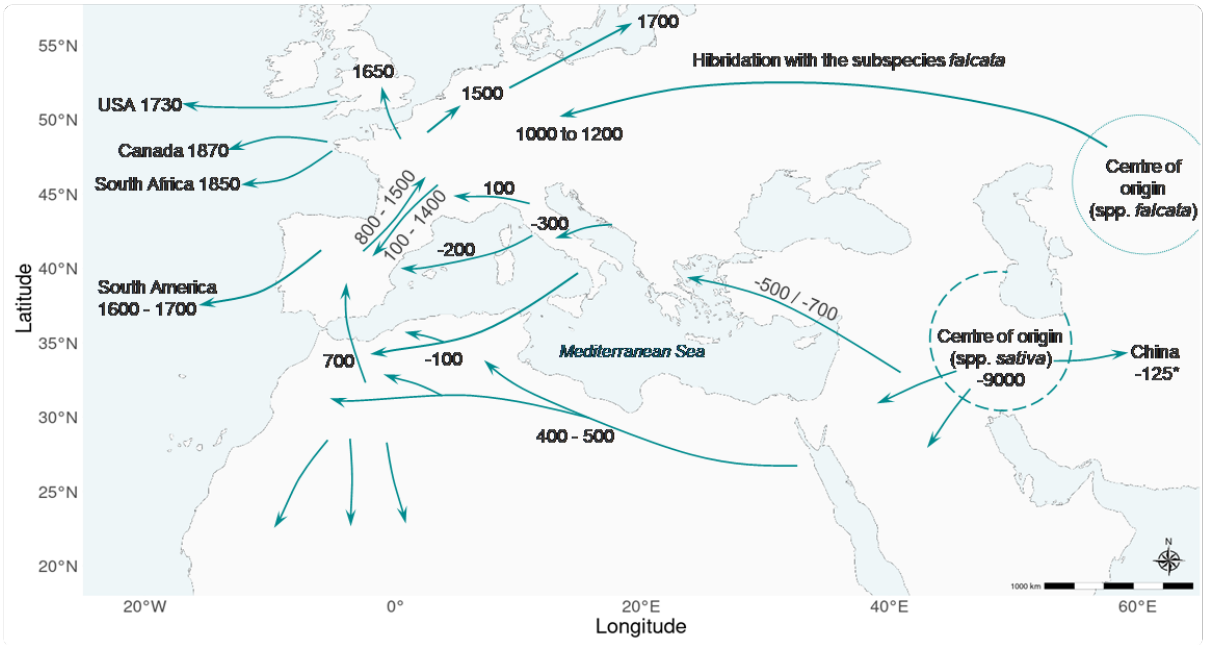


Figure 1: History of alfalfa expansion from its presumed center of origin. Modified from Prosperi et al. (2014) by Irving Arcía.

M. sativa exhibits intrinsic diversity, which is amplified by a wide spectrum of phenotypic variations resulting from processes such as hybridization, polyploidy, and domestication.

Among the taxa described in the complex, we find *sativa*, *falcata*, *glomerata*, and x *varia*. Each taxon has been considered as species or subspecies by different investigators over time (Şakiroğlu & İlhan, 2021). However, there are two main taxonomical groups: *sativa* and *falcata*. Morphologically, these subspecies are distinct (Figure 2); ssp. *sativa* features a taproot system, erect growth habit, purple flowers, and spiral pods, while ssp. *falcata* exhibits a fasciculate root system, prostrate growth habit, yellow flowers, and fasciculate pods (Teuber & Brick, 1988). *Medicago* x *varia* is commonly encountered as a result of hybridization between subsp. *sativa* and subsp. *falcata* (Prosperi et al., 2014). For clarity, in this document, the aforementioned taxa are all considered as subspecies of the *M. sativa* complex.

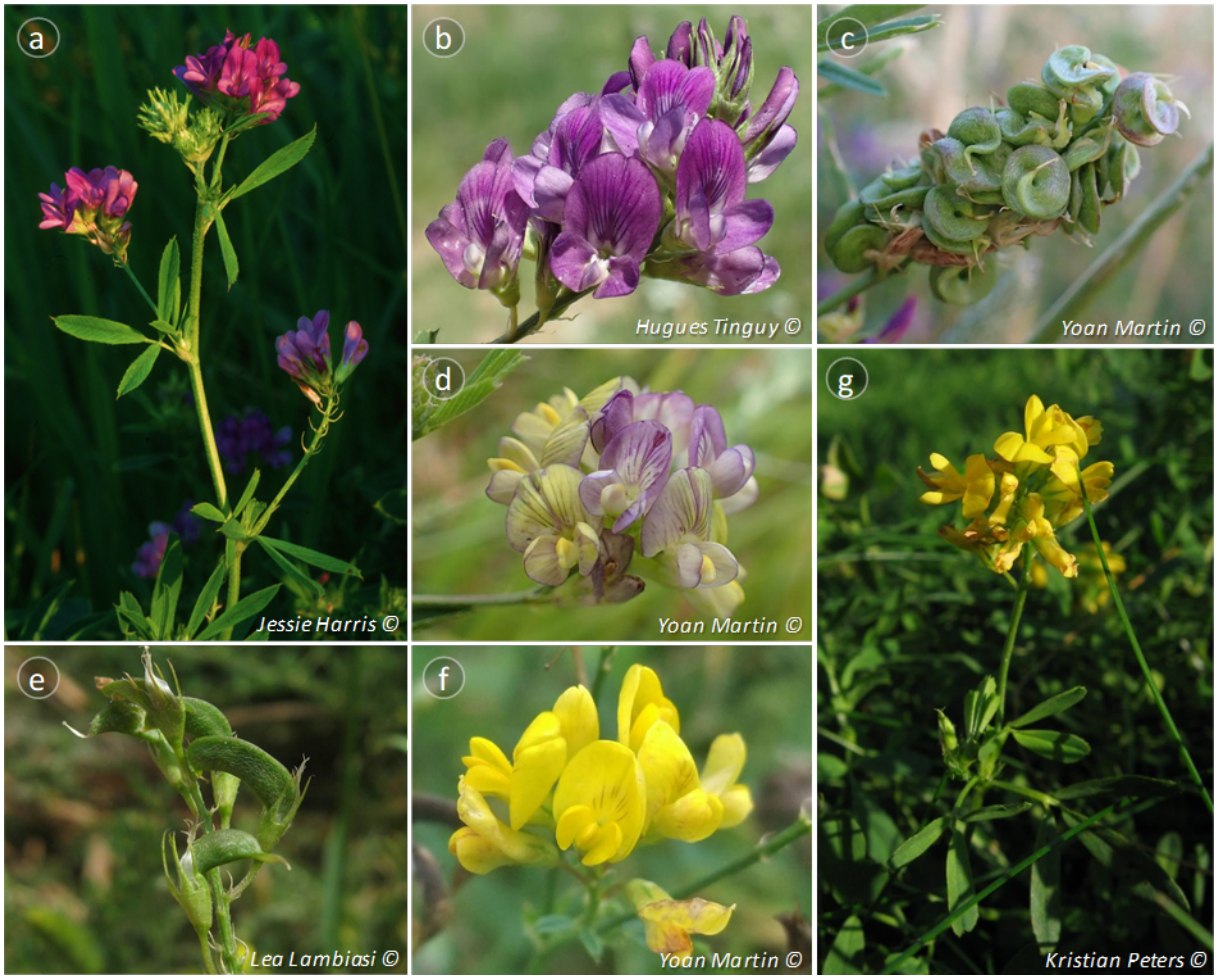


Figure 2: Main subspecies of the *M. sativa* complex. (a) Upright habit of spp. *sativa*, (b) Inflorescence of ssp. *sativa*, (c) Infructescence of spp. *sativa*, (d) Inflorescence of *M.* × *varia*, (e) Infructescence of spp. *falcata*, (f) Inflorescence of ssp. *falcata*, (g) Habit of spp. *falcata*. Modified by Irving Arcía from MNHN & OFB [Ed] (2024).

1.2.3 Breeding

Plant breeding in alfalfa is complex due to the plant's allogamous nature, which prevents the creation of pure lines and results in varieties existing as synthetic populations. On the other hand, alfalfa's perennial characteristics and long breeding cycles mean that advancements in yield and other traits occur slowly (Julier et al., 2017). Historically, the development of alfalfa cultivars has primarily utilized phenotypic recurrent selection within natural field conditions (Bolaños-Aguilar, 2001; Li & Brummer, 2012). Recurrent selection plays a critical role in this process, aiming to incrementally increase the presence of beneficial alleles while preserving genetic diversity for ongoing improvement efforts. The challenges inherent in alfalfa genetic improvement set the stage for innovative approaches in plant breeding. As traditional methods such as recurrent selection strive to incrementally enhance alfalfa's genetic traits, there emerges a clear need for advanced techniques that can accelerate and refine this process (Li & Brummer, 2012). Among those techniques we find Marker Assisted Selection (MAS)

and Genomic Selection.

1.2.4 Genetics

The genetic composition of *Medicago sativa* reflects its complex taxonomic structure. While diploids are present, *Medicago sativa* ssp. *sativa* is predominantly tetraploid, featuring a chromosome count of 2n = 4x = 32. In contrast, *Medicago sativa* ssp. *falcata* may exhibit either diploid or tetraploid forms (Quiros & Bauchan, 1988). This genetic diversity is critical for alfalfa cultivation, with tetraploid varieties generally preferred for their robustness and vigor in agricultural settings. The genome size of these autotetraploid populations ranges from 800-1,000 Mb for 1 C (Medina et al., 2020).

Alfalfa's genetic variability is primarily intra-varietal (Flajoulot et al., 2005; Julier et al., 2000), influenced by its mode of reproduction as a cross-pollinated species and its autotetraploid nature. This variability is also a result of its recent domestication and frequent seed exchanges. Regional genetic variations have been noted, with studies highlighting distinct differences in genetic diversity between Asian and Middle Eastern accessions (Muller et al., 2003), and between Chinese germplasm and that from other regions (Pégard et al., 2023; Qiang et al., 2015). It is from this germplasm that a reference genome for alfalfa has been produced (Chen et al., 2020).

The adoption of Genotyping-by-Sequencing (GBS) technology has significantly advanced alfalfa genotyping. Early efforts relied on *de novo* assembly of reads without a reference genome (Annicchiarico et al., 2015), while other studies used the genome of *M. truncatula* as a reference (Julier et al., 2018). However, because these two legumes are different species, few functional genes from *M. truncatula* have been found in alfalfa (He et al., 2022). More accurate alignments have become possible with the development of alfalfa's own reference genomes (Chen et al., 2020; Long et al., 2022; Shen et al., 2020). Recently, Pégard et al. (2023), performed GWAS and Genomic Prediction using as reference genome the one assembled by Chen et al. (2020); most of the pipeline used here follows this study.

1.3 Association Studies

Association studies, particularly genome-wide association studies (GWAS), play a pivotal role in identifying genetic variants that influence traits. GWAS focus on detecting single-nucleotide polymorphisms (SNPs). Thanks to the certain SNPs that are linked to phenotypic traits, researchers can suggest that nearby genetic variants may be crucial for these traits (Flint-Garcia et al., 2003). This process aids in the detection of quantitative trait loci (QTL), where significant QTL signals indicate a strong statistical linkage to key phenotypic traits (Flint-Garcia et al., 2003).

1.3.1 Genetic structure

Genetic similarity (or relatedness) complicates the detection of causal variants in standard association studies, often leading to the identification of numerous false positive associations (Flint-Garcia et al., 2003; Segura et al., 2012). Sul et al. (2018) identifies two main types of relatedness that contribute to high rates of false positives: differences in ancestry and cryptic relatedness. These differences can involve varying backgrounds among study participants, with

large population cohorts (N $\geq$ 5,000) frequently displaying shared ancestry across different populations. Cryptic relatedness occurs when individuals are closely related in ways unknown to the researchers, which, along with ancestry differences, constitutes population structure.

Linkage disequilibrium (LD) is another factor affecting GWAS results. Associations of a trait with multiple SNPs within a region may be due to long LD with an untyped causal variant rather than allelic heterogeneity. This issue is exacerbated in related populations, where population structure causes genome-wide LD between physically unlinked loci, leading to statistical confounding effects in genome-wide association studies (Segura et al., 2012). Furthermore, the extent of LD determines the number and density of markers needed (Flint-Garcia et al., 2003); when LD persists over longer distances, fewer markers are needed to capture the same genetic variation over a larger region of the genome. Flint-Garcia et al. (2003) also notes that selection enhances LD, creating locus-specific bottlenecks and promoting LD between the selected allele at a locus and its neighboring loci. Moreover, selection targeting a phenotype influenced by two unlinked loci can also lead to LD, despite these loci not being physically connected. As an allogamous species, alfalfa typically exhibits LD over short physical distances (Herrmann et al., 2010).

1.4 Challenges

Building on the foundation laid by existing genome-wide association studies (He et al., 2022; Medina et al., 2020; Pégard et al., 2023), there is a compelling need to extend these studies to address unique challenges associated with using alfalfa as a living mulch.

Alfalfa's role as a forage crop makes commercial varieties vigorous but competitive for resources when co-cultivated with cash crops; this can ultimately diminish the cash crop yields. To prevent competition for light, mowing or the usage of herbicides becomes necessary (Carof, 2006; Labreuche et al., 2017). Employing dormant varieties could be a strategy to let the primary crop establish; however, this also makes room for weeds to thrive. A potential compromise could involve the development of semi-dormant varieties. Consequently, there is a need to identify traits for the living-mulch alfalfa ideotype within a broad germplasm.

High seed production is a critical trait to consider. Indeed, the ability of a cultivar to provide high seed yield determines its commercial success and wide acceptance by farmers (Bolaños-Aguilar et al., 2000). This is particularly crucial for emerging alfalfa varieties intended for use as living-mulch. Moreover, it is essential to explore underutilized germplasm with unique characteristics such as shorter or erect growth habits, in particular those not previously selected for seed production. This approach could provide new insights and broaden the genetic diversity for more effective living-mulch applications.

1.5 Seed Production

In alfalfa, seed yield is determined by a hierarchy of levels extending from the broadest scale of plant spacing in the field (stand density) up to the scale of the seed itself (Bolaños-Aguilar, 2001): stand density, plant, stem, inflorescence, pod, and seed; see Figure 3. Each level of this hierarchy can be quantified either as counts or as weights, each dependent on a broader level (e.g., number of seeds per pod, number of pods per inflorescence, number of inflorescences per stem, ending by number of healthy plants per stand). This hierarchical quantification allows

determination of traits or components that contribute to the overall seed yield difference between varieties. The correlation between seed yield and the components varies; however, variables related to the inflorescence are particularly linked to seed production. Bolaños-Aguilar et al. (2000) and Bouton and Gates (2003) found that the inflorescence plays a pivotal role in seed production-related traits, which is why this work places major emphasis on traits associated with the inflorescence.

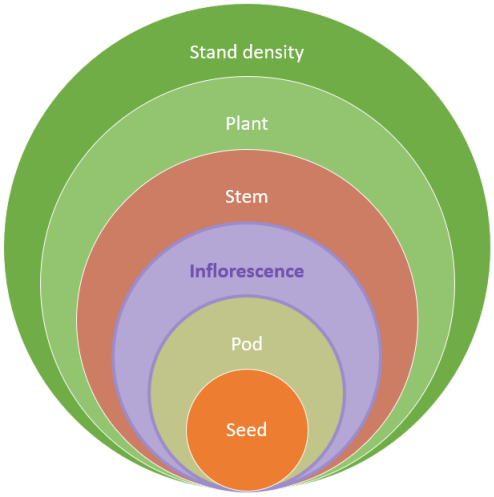


Figure 3: Alfalfa seed yield hierarchy of levels.

1.6 Objectives

The primary aim of this research is to **understand the genetic determinism of seed production in a diversified alfalfa panel meant to be used as living mulch**. To achieve this, the study is structured around several specific objectives. Initially, we will **evaluate genetic diversity**. Subsequently, we will **quantify the heritability of these traits** to determine the extent to which they are genetically controlled versus influenced by environmental factors. The research will then progress to a Genome-Wide Association Study (GWAS), which will **identify the genetic markers closely associated with seed production**.

2 Materials and Methods

2.1 Plant material

The 30 populations used in this study were derived from 22 synthetic populations and 8 wild populations (named P30, see Table 1), each represented by 40 plants, totaling 1200 individuals. In addition to the standard 40 plants per population, 155 individuals of the Milky Max variety were included to monitor spatial homogeneity. The chosen populations aimed to represent a diverse range of characteristics, including subspecies, ploidy levels, autumn dormancy, and whether they were wild or cultivated. The tetraploid populations from P30 were named P27.

Additionally, a polycross comprising 245 individuals was available (named F1). This polycross resulted from the outcrossing of three populations: Krasnokutskaya (subspecies *falcata*

), Mezzo (a vigorous cultivated variety of subspecies *sativa*), and LPIII (an experimental variety of subspecies sativa known for high seed production potential), each represented by 30 plants. In total, the study included 1600 plants.

Table 1: Overview of Alfalfa Accessions by Subspecies, Ploidy, Cultivation Status, and Origin.

Population	Subspecies	Ploidy	Type	Origin
Camporegio	falcata	4x	Cultivated	Italy
Coussouls	sativa	4x	Cultivated	France
Gabès	sativa	4x	Cultivated	Tunisia
Glomerata	glomerata	2x	Wild	France
Greenmed	sativa	4x	Cultivated	Spain
Koalf 2-96	sativa	4x	Cultivated	Hungary
Krasnokutskaya	falcata	4x	Cultivated	Ukraine
L4332	sativa	4x	Cultivated	France
L5323	× varia	4x	Cultivated	France
L8988	falcata	4x	Cultivated	France
Limory	sativa	4x	Cultivated	France
Ludelis	sativa	4x	Cultivated	France
Luzelle	sativa	4x	Cultivated	France
Malzeville	falcata	4x	Wild	France
Marais de Luçon	sativa	4x	Cultivated	France
Maron	falcata	4x	Wild	France
Mezzo	sativa	4x	Cultivated	France
Miechowska	sativa	4x	Cultivated	Poland
Milfeuil	sativa	4x	Cultivated	France
Milky Max	sativa	4x	Cultivated	France
Monte Oscuro	sativa	4x	Wild	Spain
Occitane	sativa	4x	Cultivated	France
Pancrudo	sativa	4x	Wild	Spain
Quasifalcata	falcata	2x	Wild	Russia
Romanica	falcata	2x	Wild	Russia
Speeda	sativa	4x	Cultivated	France
Timbale	sativa	4x	Cultivated	France
Verdor	sativa	4x	Cultivated	France
Villanueva de Jara	sativa	4x	Wild	Spain
Wugong	sativa	4x	Cultivated	China

2.2 Experimental Conditions

Plants were cultivated under greenhouse conditions until they reached one month of age. In April 2021, the seedlings were transplanted into the field in the experimental site illustrated in Figure 4. Each plant was spaced 70 cm apart from its neighbors. The experiment included four replicates, organized into four blocks. Plants were cultivated at the Grasslands and Fodder

Plants Research Unit of INRAE (Lusignan, France; 46° 24' 3" N, 0° 4' 53" E). Data were collected over 2022 and 2023.

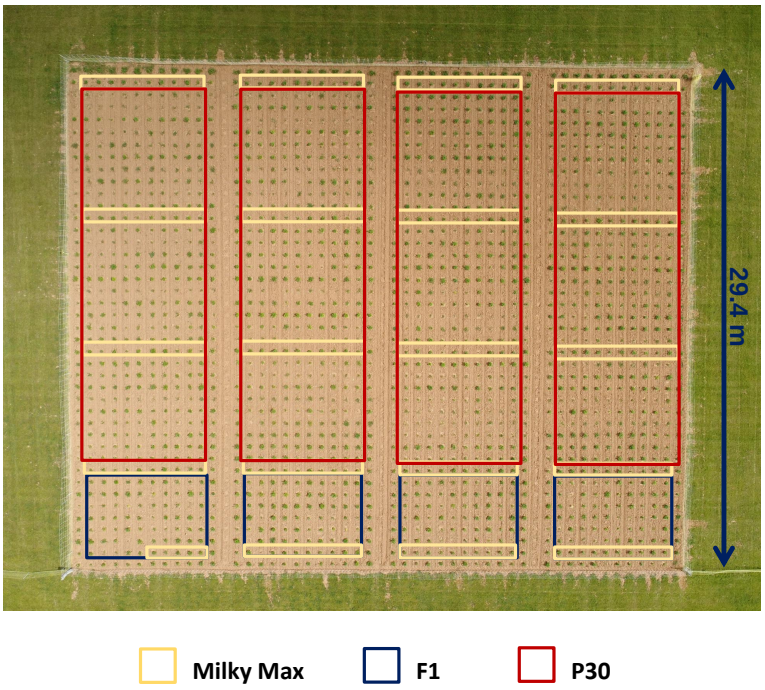


Figure 4: Spatial Arrangement of the Alfalfa Populations. This aerial photograph illustrates the layout of the populations distributed across four experimental blocks in a field study. Each block systematically hosts distinct varieties: 'Milky Max' for spatial monitoring (yellow), 'F1' (blue), and 'P30' (red). Notably, each block is surrounded by border plants to minimize edge effects and ensure consistent experimental conditions.

2.3 Phenotyping

2.3.1 Plant measurement

The phenotypic measurements were conducted in July 2022 and 2023. For each plant, five (2022) or ten (2023) large inflorescences were collected and four seed yield measurements components were recorded: the number of pods, the number of seeds, the weight of the seeds, and the weight of the inflorescences. These data were utilized to calculate the traits analyzed throughout this study (Table 2). In 2022 and 2023, 313 and 364 plants, respectively, could not be phenotyped. The Interquartile Range (IQR) method was used to flagged outliers, that were subsequently removed. Besides outliers' removal, after contextual and visual examination, anomalous points were also removed. Although only data points were removed at each time, sometimes this led to the removal of all variables for a plant, in total 1279 and 1216 plants were kept for 2023 and 2022 respectively.

Table 2: Overview of Seed Yield Components Measured in This Study. Each variable is classified under three main levels: Inflorescence, Pod, and Seed.

Level	Description	Variable
Inflorescence	Seed weight per inflorescence	SWpI
	Seed number per inflorescence	SNpI
	Pod number per inflorescence	PNpI
	Inflorescence weight	IW
Pod	Seed weight per pod	SWpP
	Seed number per pod	SNpP
Seed	Thousand seed weight	TSW

2.3.2 Statistical Analysis

To assess variability among accessions, years, and experimental blocks, an ANOVA was performed on P30 using the model shown in Equation 1. Here, Y_{ijk} is the measurement of the trait, m is the overall mean, a_i represents the fixed effect of the i-th accession, y_j represents the fixed effect of the j-th year, $(ay)_{ij}$ is the interaction between the accession and the year, b_k is the random effect associated with the k-th block, e_{ijk} is the residual error.

$$Y_{ijk} = m + a_i + y_j + (ay)_{ij} + b_k + e_{ijk} \tag{1}$$

To control for the potential impact of different years, separate ANOVAs were performed for each year using the model shown in Equation 2. This approach ensures that the fixed effect of the year, a_j, and its interaction with the variety, $(ta)_{ij}$, do not influence the assessment of variety and block effects within each year. The model for the yearly analysis was:

$$Y_{ik} = m + a_i + b_k + e_{ik} \tag{2}$$

2.4 Genotyping

For each plant, leaflet sampling and DNA extraction were performed according to the method described by Julier et al. (2018). Subsequently, six GBS libraries were constructed with DNA being digested by PstI and MseI enzymes based on the approach outlined by Elshire et al. (2011). Following adaptor ligation and PCR amplification, the process includes cleanup steps and final library assessment for sequencing readiness. Sequencing was done with Novaseq technology.

2.4.1 SNP calling

SNP calling followed the methodology detailed by Pégard et al. (2023), reads were mapped to chromosome B of the *Medicago sativa* reference genome (Chen et al., 2020). The analysis was restricted to biallelic SNPs and was focused on calculating SNP frequencies using custom scripts, rather than determining SNP dosages.

2.4.2 GBS Quality Control

The SNP data mapped onto scaffolds were first excluded. We then applied a filter to remove plants exhibiting over 60% missing data. Subsequently, SNPs with more than 20% missing data were removed, along with SNPs where the minor allele frequency (MAF) was below 0.05. After these filtering steps, we created matrix M1, which retained only the markers that had no missing data; resulting in a total of 13,883 markers. For the creation of matrix M2, imputation was conducted; for each variety, the MAF was calculated and used to impute missing data. Markers absent in any of the varieties (totaling 35) were excluded, culminating in matrix M2 containing 62,150 markers.

2.5 Population Structure

The M1 GBS matrix was utilized for the analysis of genetic structure. Linkage disequilibrium (LD) was measured by the partial correlation coefficient between each pair of markers Lin et al. (2012), and was plotted against the physical distance. LD was estimated for F1 (216 individuals), Milky Max (178 individuals), P27 and P30. Additionally, a Principal Component Analysis (PCA) was conducted on P30 and F1 using the R package FactoMineR (Lê et al., 2008) to assess diversity and structure within the population.

2.6 Genetic Model and Phenotypic Adjustment

2.6.1 Spatial Adjustment

When phenotypic data values distribution is structured and influenced by varying environmental conditions across the test field, a microenvironmental adjustment is necessary to correct for spatial heterogeneity bias. While doing this correction, the genotypic variance (by means of a Genetic Relationship Matrix, GRM [see Equation 4]) and the fixed block effects were also considered. The function remlf90 from the breedR package (Muñoz & Sánchez, 2024) was used to perform the correction. This package uses the Restricted Maximum Likelihood (REML) statistical method to capture the **variance components**; moreover, to model small-scale environmental variations, the tensor product of two B-splines bases was used to create a smooth representation of spatial effects. A covariance structure is used to determine how random effects (in this case, Random Knot Effects, RKE) correlate with each other across the two dimensions (rows and columns). As mentioned by Pégard et al. (2023), breedR optimized the knot numbers by an automated grid search based on the Akaike Information Criterion (AIC). Besides variance components, the function remlf90 also allows us to estimate the genetic correlation between different traits. The previous variances and effects (calculated following the model presented in equation 3) determine the microenvironmental effect, which is subtracted from the observed phenotypic value to obtain an adjusted phenotypic value. This adjustment was done only for P27; for populations F1 and Milky Max the reml90 model did not capture spacial heterogeneity.

2.6.2 Genetic Model

In this work, a mixed (both fixed effects and random effects) genetic model was employed. The phenotypic values, y, are modeled as the sum of the global mean, μ, the genetic effects, Xu, the spatial effects, Ws; and the residual effects ϵ:

$$y = \mu + \mathbf{Xu} + \mathbf{Ws} + \epsilon \tag{3}$$

While X is an incidence matrix that links each plant to its genetic effects, u is a vector representing the genetic effects. These genetic effects are assumed to be normally distributed with a mean of zero and variance $G\sigma_a^2$, where G is the genomic relationship matrix (detailed below, Equation 4), that quantifies genetic similarities between the plants, and σ_a^2 is the additive genetic variance. Ws is the spatial effects that might affect the phenotype, like variations within the field where the plants are grown. W is also an incidence matrix linking these effects to each plot, and s is a vector of these random spatial effects, which are also assumed to follow a normal distribution with zero mean and specific variance ($S\sigma^2$), S describes how these spatial effects are related across different plots. ϵ represents the residual effects (or error) that cannot be explained by genetic or known spatial factors. It is assumed to follow a normal distribution with zero mean and its own variance σ_e^2.

2.6.3 Genetic Relationship Matrix Estimation

GRM was calculated on M1 according to VanRaden (2008) and modified as presented by Cericola et al. (2018) to deal with allelic frequencies:

$$GRM = \frac{MM'}{\frac{1}{n}\sum_{j=1}^{m} \hat{p}(1-\hat{p}_j)} \tag{4}$$

With M representing the allelic frequencies matrix centered by minor allele frequency (MAF), m the number of markers, $\hat{p}_j$ the frequency of the j^{th} marker, and n a scalar meant to obtain a diagonal mean of 1, Cericola et al. (2018) defines it as the ploidy number. In this work, setting $n = 4$ achieved a diagonal mean close to 1 for tetraploid P27. Including other populations altered the diagonal mean, and using a different ploidy number lacked biological sense. Consequently, the reml90 model was not run for P30.

2.6.4 Heritability Estimation

Variance components were autonomously computed from adjusted phenotypic values using the reml90 function from the breedR package, without the need for manual supervision. This calculation adhered to the genetic model previously described (Equation 3). With these variance components, the broad-sense heritability, denoted as H^2, was determined using the following equation:

$$H^2 = \frac{Var_G}{Var_G + Var_E + Var_R} \tag{5}$$

With Var_G accounting for the genetic variance, Var_E the microenvironment effect variance, and Var_R the residual. H^2, was estimated for P27 across two different years (Table **??**), as well as for each year —done by adding the year affect to the model presented in Equation

3 . It was also calculated for F1 and Milky Max variety using (Table **??**) data from 2023, using AR1XAR1 instead of B-splines, which may result in a less accurate representation of the spacial effect; in these cases, no spatial correction was done because the model failed to identify spatial structure.

2.7 GWAS

Using 62,150 markers Genome-Wide Association Studies (GWAS) were conducted on P27 (one study per year and a third averaging data from both years), MilkyMax-2023, and F1-2023 to identify potential QTLs, this time M2 matrix was used. To mitigate bias from population structure and variation due to kinship among accessions, a multi-locus mixed model (MLMM) approach was employed using the R package mlmm, developed by Segura et al. (2012). The model was parameterized to run up to ten (non-inclusive) steps. A Bonferroni corrected threshold of 0.05 over the total number of markers was used.

3 Results and Discussion

3.1 Phenotypic data

An essential aspect of this study was to determine whether data from both years could be integrated effectively. Utilizing the relm90 model in breedR, the year effect emerged as the predominant factor for all traits. Likewise, analysis through an ANOVA (Table 3) revealed significant interactions between variety and year for all traits ($p < 0.001$), except for TSW, which had a p-value of 0.65. This could be due to environmental variations or other annual factors influencing the traits. Furthermore, while IW shows an exceptionally high F value for the year effect —suggesting a very strong year dependency, TSW exhibits no significant interaction between variety and year, indicating that the year-to-year variation in variety effects is relatively consistent for this trait compared to others.

Table 3: Refined ANOVA P-values for different variables and effects, showing <0.001 where applicable.

Variable	Variety	Year	Block	Year-Variety Interaction
PNpl	<0.001	<0.001	<0.001	<0.001
IW	<0.001	<0.001	0.00949	<0.001
SNpl	<0.001	<0.001	<0.001	<0.001
SWpl	<0.001	<0.001	0.00319	<0.001
SWpP	<0.001	<0.001	0.0218	<0.001
SNpP	<0.001	<0.001	0.109	<0.001
TSW	<0.001	0.426285	<0.001	0.650465

When run separately (Appendices 3 and 4), for both years ANOVA results indicated strong influence of variety for all variables ($p < 0.001$). This is a consistent pattern across years and suggests that genetic differences among varieties are important determinants of the traits.

The block effect is also significant for most variables, indicating that environmental or spatial variations within the blocks are affecting the traits.

3.2 Genotypic data

The percentage of markers mapped onto scaffolds corresponded to 10% of the markers. After filtering steps, the total number of markers was of 62,150 SNPs (Table 4).

Table 4: Processing Steps and SNP Counts

Step	**SNP count**	**Individual Plant Count**
Initial count data	173,876	1,557
Removing duplicated plants	173,876	1,544
Scaffolds filter	155,818	1,544
Less than 60% NA in plants	155,818	1,505
Less than 20% NA in SNPs	63,649	1,505
MAF < 0.05	62,185	1,505
SNPs present in all populations	62,150	1,505

SNPs were not evenly distributed along the chromosomes (Figure 5), and there were areas with less coverage, particularly towards the central parts of some chromosomes, possibly due to centromeric regions (Medina et al., 2020). Interestingly, in a similar way to Medina et al. (2020) and Pégard et al. (2023), chromosome 6 are among the chromosomes with the least presence of markers, this is due to a less density in the presence of coding genes (Chen et al., 2020).

Figure 5: Distribution of GBS SNP markers across the eight alfalfa chromosomes using a 500kb window

3.3 Population Structure

3.3.1 Principal Component Analysis

The PCA analysis conducted on the entire population panel (Figure 6) revealed distinct clustering patterns. Diploid varieties were markedly separated from tetraploid varieties. Additionally, the French diploid Glomerata population formed a distinct cluster, separate from the Russian diploids Romanica and Quasifalcata. The genetic divergence observed in *glomerata*, compared to other groups, supports the argument for classifying this group as a separate species, as suggested by some authors (Şakiroğlu & İlhan, 2021).

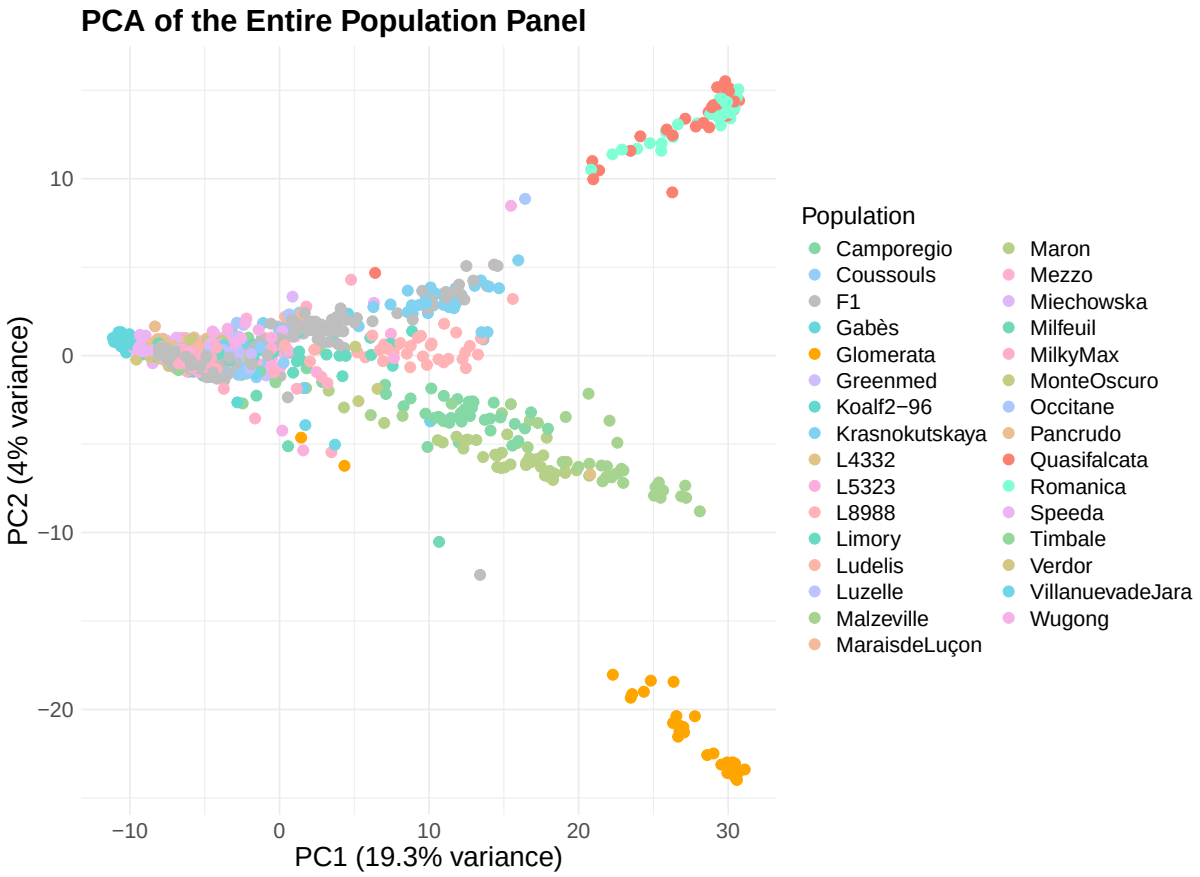


Figure 6: PCA for P30 populations.

When PCA was performed on P27, excluding the diploids (refer to Appendix 5), the data exhibited less clustering, but still some can been seen; for example for cultivated populations, this can be due to breeding selection pressure. Spanish wild populations (Pandcrudo, Villanueva de Jara and Monte Oscuro) also appear to have some clustering. F1 individuals appeared more or less centered with respect to other populations.

However, within the F1 group, there was more internal structuring (Figure 7). Depending on the maternal heritage, it was evident that out-crosses occurred, along with in-crosses between plants of the same variety and/or possible selfing within a single plant. Furthermore, there might be more intercrossing between *sativa* maternal populations (LPIII and Mezzo). This apparent preferential intercrossing may be due to differences in the developmental stage; for instance, LPIII and Mezzo may have flowered at the same time, and earlier or later than Krasnokutskaya. In this study we neither establish a difference between selfing and crosses between plants of the same variety, nor determine the paternal heritage status of the F1 plants.

However, within the F1 group, distinct internal structuring was observed, as illustrated in Figure 7. This structuring highlights the occurrence of both out-crosses and in-crosses among plants of the same variety, as well as potential selfing within individual plants. Notably —regardless of the random distribution of plants in the field, there appears to be a higher incidence of intercrossing among the *sativa* maternal populations (LPIII and Mezzo), possibly influenced by differences in their developmental stages; for example, LPIII and Mezzo may have matured earlier or later than Krasnokutskaya. Differences in inflorescence architecture

and seed viability may play a significat role too. In this study, a differentiation between selfing and in-crosses within the same variety is not done, nor was it possible to determine the paternal heritage of the F1 plants. Further genetic characterization is needed to elucidate the genealogical history of the F1 population.

In the context of associations studies, including diploid varieties may increase the risk of producing population stratification bias. This is because genetic differences between groups might be associated with both the genetics and the phenotype, potentially leading to false associations; associations may represent adaptations to specific environmental or historical contexts. Another source for bias may be due to the high relatedness of F1. To avoid this bias, separated analysis were done for P27 and F1 when calculating the component variances and performing the GWAS.

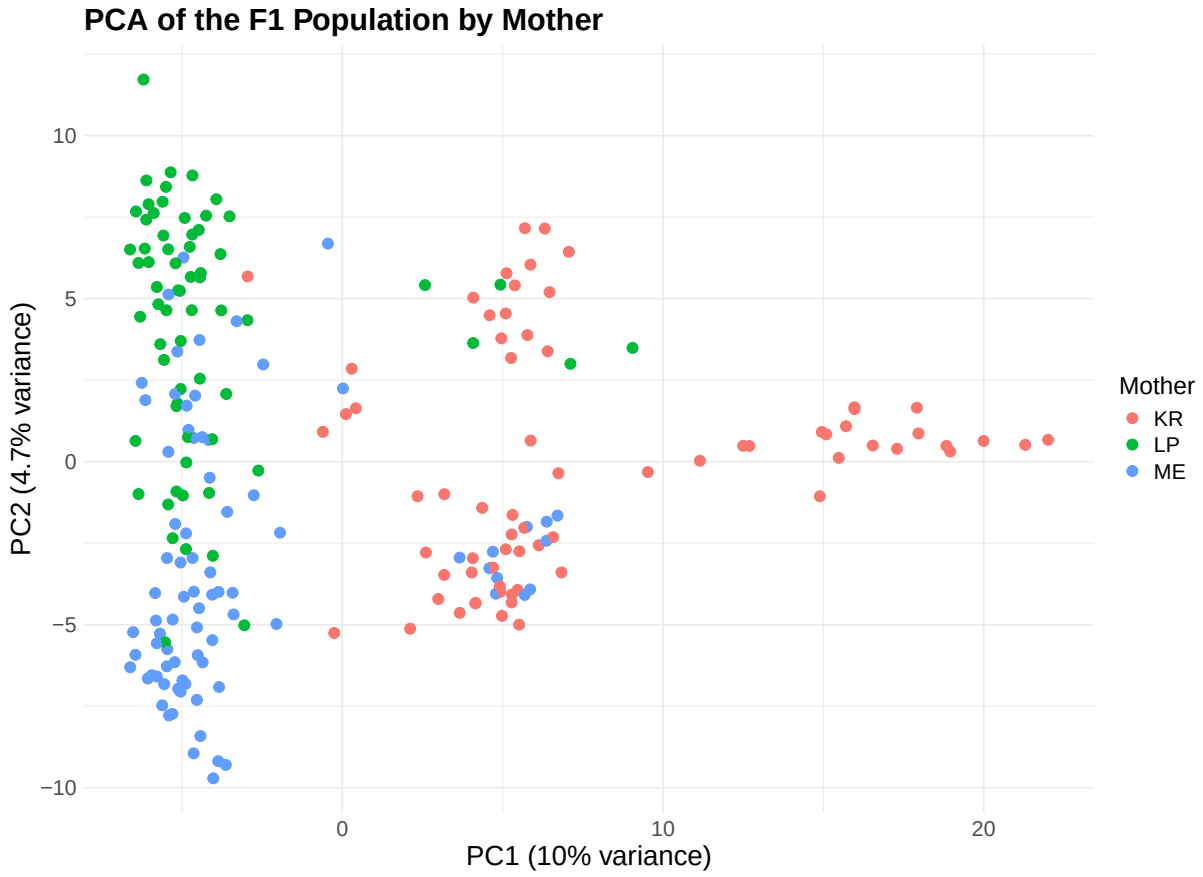


Figure 7: PCA for F1 population. The plot shows the distribution of individuals based on their genetic variation, with each color representing the maternal heritage of the individuals: ME for Mezzo, KR for Krasnokutskaya, and LP for LPIII.

3.3.2 Linkage Disequilibrium

Generally, all populations exhibit an expected LD decay where LD decreases as physical distance between SNPs increases. This decay pattern is a hallmark of recombination breaking down linkage between loci over generations. The LD decay in P27 and P30 was relatively steep (Figure 8), with high LD values concentrated at short distances and rapidly decreasing over

longer distances. This steep decay might be indicator of populations with high recombination rates and substantial genetic diversity. The P30 population, which included the P27 set plus three additional diploid populations, showed a broader distribution of LD values. The inclusion of additional populations appeared to increase the variability in LD, with some SNP pairs maintaining higher LD over longer distances. This can be attributed to the increased genetic variability and potential introduction of new haplotypes from the additional diploid populations.

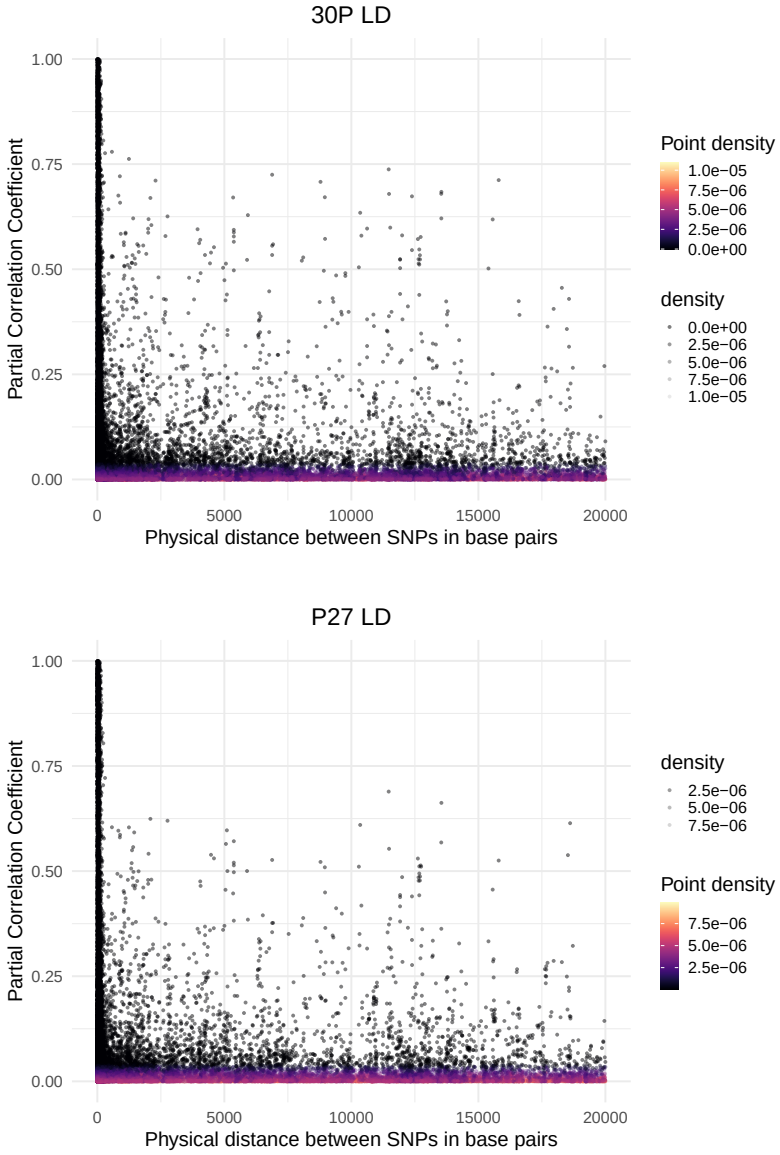


Figure 8: Linkage Disequilibrium plots for P30 (top) and P27 (bottom).

The F1 population also shows a decay pattern where the LD decreases as the physical distance between SNPs increases (Figure 9). The relatively quick decay of LD suggests effective recombination events in the F1 population, which might be related to the genetic mixing from the three parental populations. In contrast to P27 and P30, F1 exhibited a slower decay, which

may be due to some chromosome fragments not having the opportunity to recombine over several generations.

Surprisingly, Milky Max displays a similar LD decay pattern to the F1 population, though with some differences in the density and spread of LD values. The initial high LD values decrease with increasing physical distance, possibly indicating recombination over generations. However, the point density suggests there might be more regions with moderate LD compared to the F1 population. Also, Milky Max population shows a less steep LD decay pattern than F1. This could be indicative of less effective recombination and genetic mixing due to the breeding history and selection pressures specific to the Milky Max, where fewer recombinations than expected occurred.

The fact that Milky Max shows a LD profile similar to that of F1, rather than P30, may be due to its breeding history. Given that the initial linkage disequilibrium (LD) in a synthetic variety is shaped by the number and genetic relatedness of its parent plants, and typically diminishes through successive panmictic generations (Gallais, 2003), several scenarios could explain the LD observed in Milky Max. It is plausible that Milky Max was developed either from a relatively small number of parents, from parents who were closely related, or both.

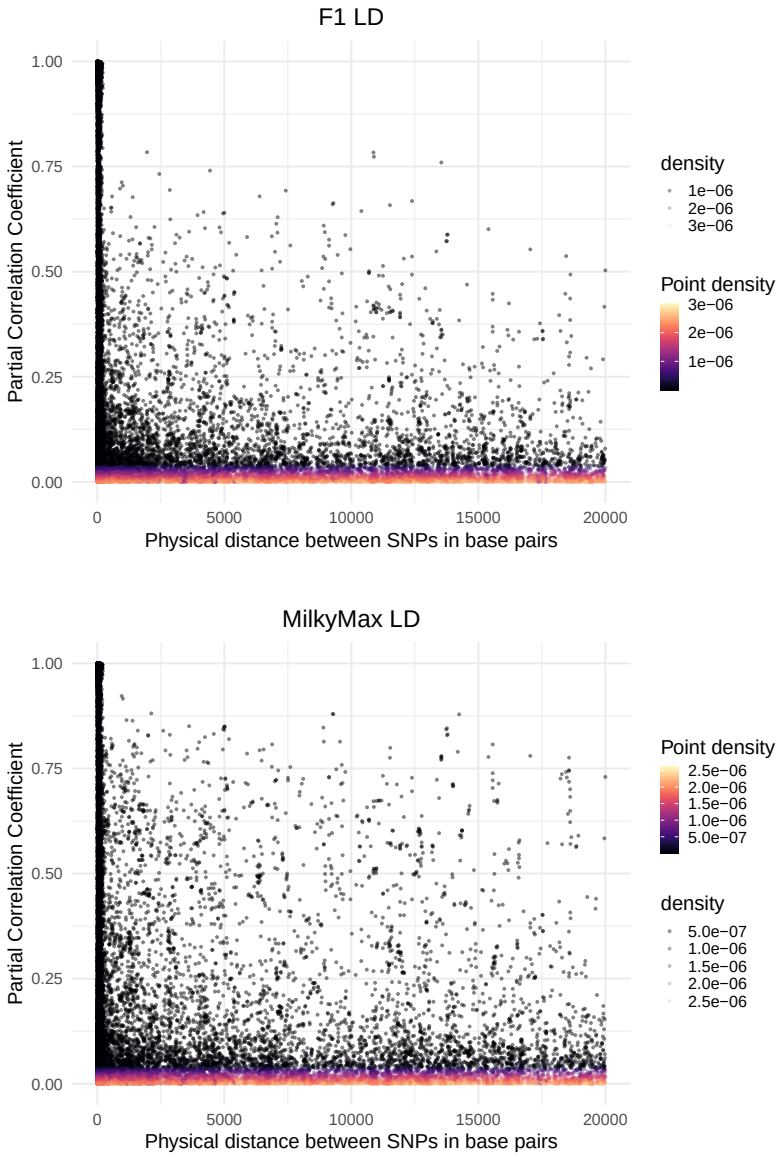


Figure 9: Linkage Disequilibrium plots for F1 (top) and Milky Max (bottom).

3.4 Heritability and Correlation

The heritability estimates for P27-2023 were higher than those for P27-2022, and also surpassed the estimates when both years were analyzed collectively. When analyzed across both years, heritability exhibits intermediate values between those observed in 2022 and 2023 (except for TSW, with the highest heritability across years).

Heritability is known to be adversely affected by unfavorable environments and has been reported to change, often improving, with plant age (Lawrence-Paul et al., 2023). Therefore, variation in heritability across years could be attributed to experimental complications encountered during data collection in 2022, where inflorescences were collected later —and in a less benign environment. Conversely, the improved heritability in 2023 might be linked to

the plants being older and more established.

Although most traits exhibit varying levels of heritability across different years for P27 (Table 5), with some showing significant changes —suggesting environmental and/or experimental impacts on trait heritability), in overall, SWpP, SNpP and TSW (when analysed each year independently) are the traits with the lowest heritability. SWpP was found to have the smallest heritability for Milky Max and P27 in both years, similarly to TSW its value increases when used both years.

For SWpI, PNpI, SNpI, SWpP and SNpP Bolaños-Aguilar et al. (2000) presents broad heritabilities of 0.42, 0.32, 0.42, 0.39, and 0.38 respectively for one experiment and 0.31, 0.27, 0.34, 0.34, and 0.40 for another. Thus, although not as wide as the heritabilities found in this work, some variation is observed depending on experimental conditions. Also, these values exhibit moderate heritabilities, which aligns with ours.

Table 5: Broad-sense heritability estimates for P27 across different years and for F1 and Milky Max in 2023.

Trait	P27			2023	
	2022	2023	2022-2023	F1	Milky Max
PNpI	0.25	0.46	0.34	0.27	0.28
IW	0.30	0.50	0.29	0.49	0.16
SNpI	0.26	0.41	0.25	0.56	0.35
SWpI	0.23	0.39	0.23	0.56	0.04
SWpP	0.16	0.16	0.28	0.51	0.22
SNpP	0.23	0.32	0.21	0.45	0.49
TSW	0.29	0.34	0.46	0.23	0.84

The correlation analysis for P27 (10) revealed a strong alignment between phenotypic and genotypic correlations among the traits. Traits associated with inflorescence (XX-pI and IW), exhibit strong positive correlations. Similarly, traits linked to pod development (XX-pP), with the exception of PNpI, also demonstrate robust positive correlations among themselves. This suggests that certain traits could potentially be predicted from others, which could simplify breeding programs or genetic studies by focusing on a subset of traits. On the other hand, similarly to Lefebvre (2023), TSW displays very weak correlations with most other traits, indicating a distinct set of influencing factors. This independence suggests that the factors driving TSW are likely different from those impacting inflorescence and pod traits, possibly involving unique genetic or environmental interactions. Similar behaviour was seen for the phenotypic correlations for Milky Max and F1 in 2023.

In general, for SWpI, PNpI, SNpI, SWpP and SNpP Bolaños-Aguilar et al. (2000) reports moderate to high positive genetic and phenotypic correlations, with the exception of PNpI which exhibits low correlations. This conforms with the correlations on Figure 10, where (after TSW) PNpI appears to be the least correlated trait.

The fact that TSW is slightly negatively correlated with SNpP and SNpI may indicate a possible negative compensation in seed weight when there is a high seed yield. This is something to keep in mind when breeding for seed yield in alfalfa, because it could increase

at the expense of seed size and weight. In this scenario, seeds may present problems for germination, since smaller seeds may establish poorly (Wall & Steppuhn, 2007).

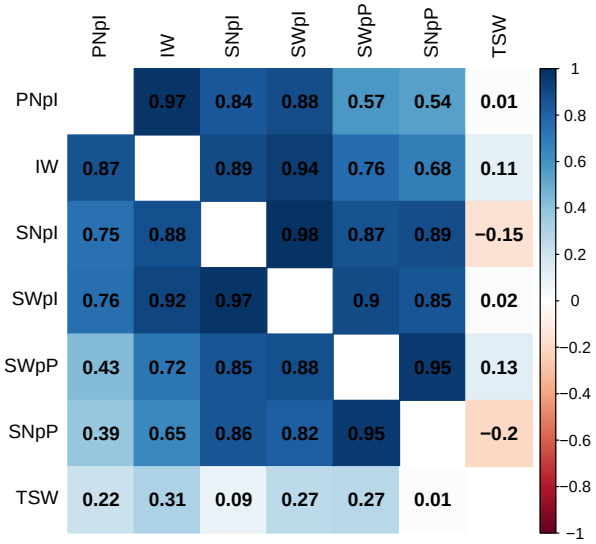


Figure 10: Correlation between traits for P27. Phenotypic and genetic correlations (based on the covariance matrix estimated with a multi-trait model) are on the left and right sides of the diagonal, respectively.

3.5 GWAS

No QTLs were identified for P27 in 2022, which aligns with expectations given the low heritability observed for all traits. However, a combined analysis for both years identified five QTLs. The r2 values, which measure the proportion of variance in the trait explained by the markers, are relatively low in some cases (see Appendix 12), suggesting either complex genetic architecture with multiple contributing factors or insufficient marker coverage. Higher r2 values in certain markers (e.g., associated with IW) indicate stronger associations worth further investigation.

Since heritability quantifies the proportion of phenotypic variance attributable to genetic variance within a specific population, high heritability can suggest the presence of multiple genetic factors, such as additive, dominance, and epistatic effects, influencing the trait (Visscher et al., 2008). This could imply a complex genetic architecture with potential for multiple QTLs. However, high heritability does not equate to genetic determinism (Visscher et al., 2008). Rather, it indicates that genetics significantly influence phenotypic variation in the current population context. Thus, in cases where GWAS fails to identify significant markers despite high heritability (e.g., SNpl and SWpl for F1 in 2023; and TSW for Milky Max in 2023), it may be due to the complex genetic architecture involving many small-effect alleles or other factors limiting detection power.

On the other hand, traits exhibiting both high genotypic and phenotypic correlations are anticipated to share QTLs (Gardner & Latta, 2007). This is the case for population P27 in 2023; the GWAS results highlighted a marker, chr1.2_52718807, which appeared to be

associated with most traits (SNpP, SNpI, SWpI, SWpP and on the second GWAS step for IW, p-value 1.91e-06), underscoring its potential significance.

Thanks to the annotated reference genome (Carrere & Gouzy, 2024), it could be found that most of the QTLs fell near or within an annotated region. Those annotations may be linked to transcriptional differences in specific metabolic pathway steps. For instance, some of the annotations for the found significant markers relate to DNA-binding or transcriptional regulation such as SANT/Myb domains. MYB transcription factors are known for their role in various plant processes including development and stress responses. Their presence in inflorescence elongation (Huang et al., 2024) and potential implication in the Inflorescence Weight (maker chr2.2_60522359 in Table 6, Appendix 11) through related regulatory domains highlight their relevance. The same marker has an annotation for a Zinc finger domain. This kind of domains are commonly associated with DNA binding and are often found in transcription factors, which are crucial for regulating gene expression. Finding this kind of domain in the our GWAS results (Table 6) suggests a role in transcription regulation that could be relevant to both transcriptional control of inflorescence elongation (Huang et al., 2024) and genetic variations affecting Inflorescence Weight.

Likewise, an annotation for a UDP-glucosyltransferase (UGT) is captured through the marker chr8.2_36886928 (Appendix 10) in association with SNpP. UGTs participate in stress responses by utilizing various phytohormones and secondary metabolites as substrates (Gharabli et al., 2023). An example occurs in rice, where one particular UGT mediates metabolic influx from lignin to flavonoid biosynthesis under abiotic stress conditions (Dong et al., 2020). Expression of this UGT upregulates flavonoid-mediated auxin levels, which in turn increases grain size. Similar pathways may influence SNpP or even seed size in alfalfa.

Table 6: Genomic Associations Identified in Alfalfa Seed Production Traits. This table summarizes markers identified through GWAS, noting their genomic locations, associated traits, and datasets used for where the QTLs were found. The last column lists proteins or domains potentially linked to these markers, providing insights into the underlying biological mechanisms influencing seed production traits. Functions/names are given according to UniProt descriptions. When no adjustment was done, *-WA* was added to the dataset name.

Marker	Traits	Dataset	Candidate Annotation
chr1.2_36542070	SWpP, SWpI	P27-22/23	NA
chr1.2_52718807	SNpI, SWpI, SWpP, SNpP	P27-23	Eisosome protein; aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit; dentin sialophosphoprotein-like
chr2.2_11993317	SNpP	MilkyMax-23WA	Casparian strip membrane protein
chr2.2_21391439	IW	F1-23WA	NA
chr2.2_60522359	IW	F1-23WA	ADA2-like, zinc finger, ZZ-type; SWIRM domain, SANT/Myb domain
chr3.2_28840119	IW	F1-23WA	Uncharacterized
chr3.2_71781907	SNpP	MilkyMax-23WA	Ankyrin repeat-containing domain superfamily
chr3.2_79337878	IW	P27-23	ZIP transporter
chr4.2_73304915	SNpP	P27-22/23	NA
chr4.2_92760839	SNpP	MilkyMax-23WA	F-box domain
chr5.2_77029282	IW	F1-23WA	Anaphase-promoting complex subunit 5 domain
chr6.2_452788	SNpI	P27-22/23	SKP1 component, POZ domain
chr6.2_84468503	TSW	P27-22/23	SLC26A/SulP transporter domain
chr8.2_11767356	SNpP	MilkyMax-23WA	NA
chr8.2_36886928	SNpP	MilkyMax-23WA	UDP-glucuronosyl/UDP-glucosyltransferase
chr8.2_72610801	TSW	F1-23WA	Pyridoxal phosphate-dependent transferase, major domain

4 Conclusions

This study aimed to elucidate the genetic blueprint of seed production in alfalfa, intended for use as living mulch, by examining a diverse alfalfa panel. Genetic determinism could be effectively addressed thanks to the evaluation of genetic diversity, heritability quantification, and

identification of genetic markers via Genome-Wide Association Studies (GWAS). Significant markers found on the Genome-Wide Association Studies provide leads for further functional studies to confirm the roles of genes linked to identified QTLs. Expanding the study to include more markers through denser sequencing methods could capture a broader genetic variance, integrating genetic insights with physiological studies to enhance breeding strategies comprehensively. Likewise, further research is required to validate these markers to determine their consistency and utility. If the markers found during the GWAS are validated as QTLs, they can guide future breeding programs (MAS) aiming to improve seed yield.

This work emphasizes the importance of considering annual variability in agricultural research and breeding strategies. Future studies should perform year-specific analyses to better tailor strategies that take advantage of or mitigate the effects of yearly environmental fluctuations, thereby improving the predictiveness and effectiveness of breeding and management practices.

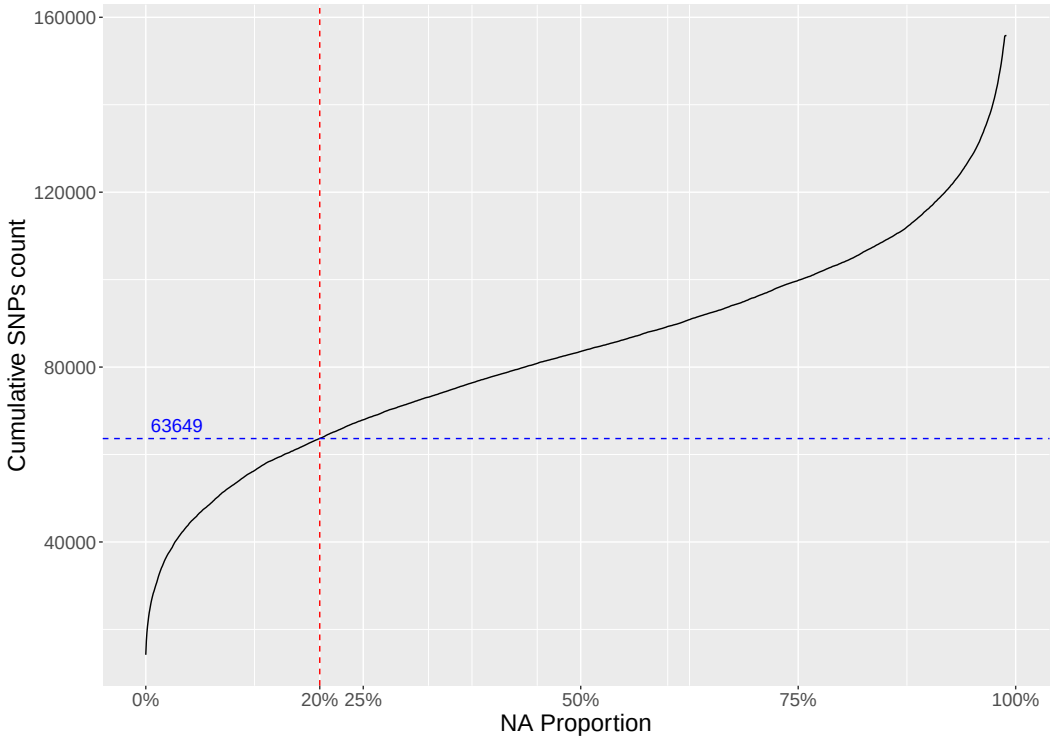
Additionally, considering the genetic markers already identified and filtered through our study, there is an opportunity to employ genomic prediction techniques. By using these markers, we can potentially improve the predictability and effectiveness of breeding strategies for alfalfa. Genomic prediction could offer a more precise approach to estimating phenotypic outcomes based on genetic data, potentially leading to better results in developing varieties with optimized seed production and living mulch qualities. This approach could further refine our understanding of the genetic determinants of alfalfa's seed yield and enhance the application of our findings in practical breeding programs.

One final remark concerns the Genomic Relationship Matrix (GRM) used during the GWAS implementation and the component variances computation. There is a need for continued research on the computation of the GRM, especially concerning the use of allelic frequencies versus allelic dosages in panels with varying ploidy levels, which is crucial for accurate genetic analysis and breeding applications.

5 Appendix

5.1 GBS Data Processing

Appendix 1: Cumulative Count of SNPs by Missing Data Proportion



5.2 Phenotypic Data Processing and Analysis

Appendix 2: Summary of Plant Data Processing Steps Across Years. The table shows the number of plants at various stages of data processing for the years 2022 and 2023, illustrating the impact of each filtering and review step.

Step	2022	2023
Initial Plant Count	1236	1287
Plants after IQR Filtering	1498	1286
Plants after Calculating Secondary Variables	1217	1280
Plants after Manual Review	1216	1279

Appendix 3: ANOVA P-values for different variables without year-variety interaction, 2022.

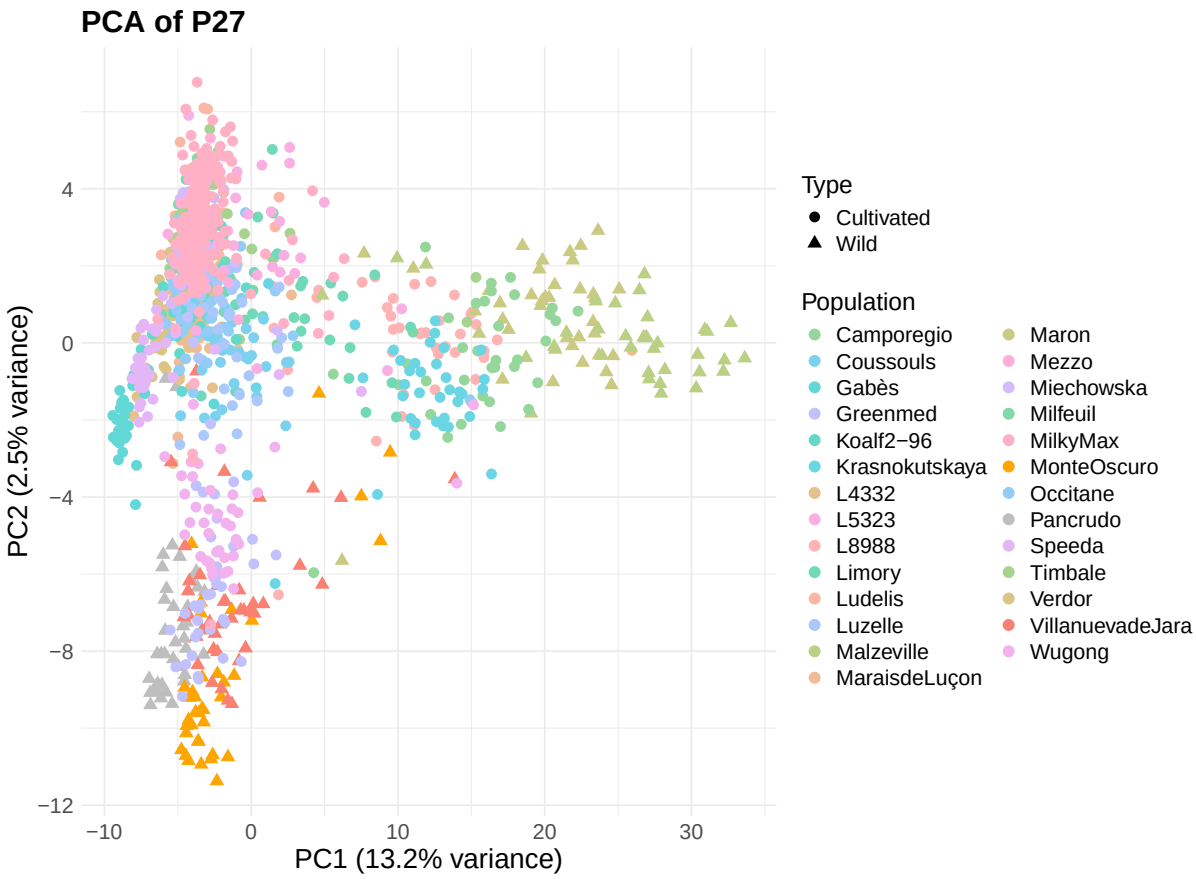
Variable	Variety (P-value)	Block (P-value)
PNpI	<0.001	<0.001
IW	<0.001	0.00784
SNpI	<0.001	<0.001
SWpI	<0.001	<0.001
SWpP	<0.001	0.0031
SNpP	<0.001	0.00179
TSW	<0.001	<0.001

Appendix 4: ANOVA P-values for different variables without year-variety interaction, 2023.

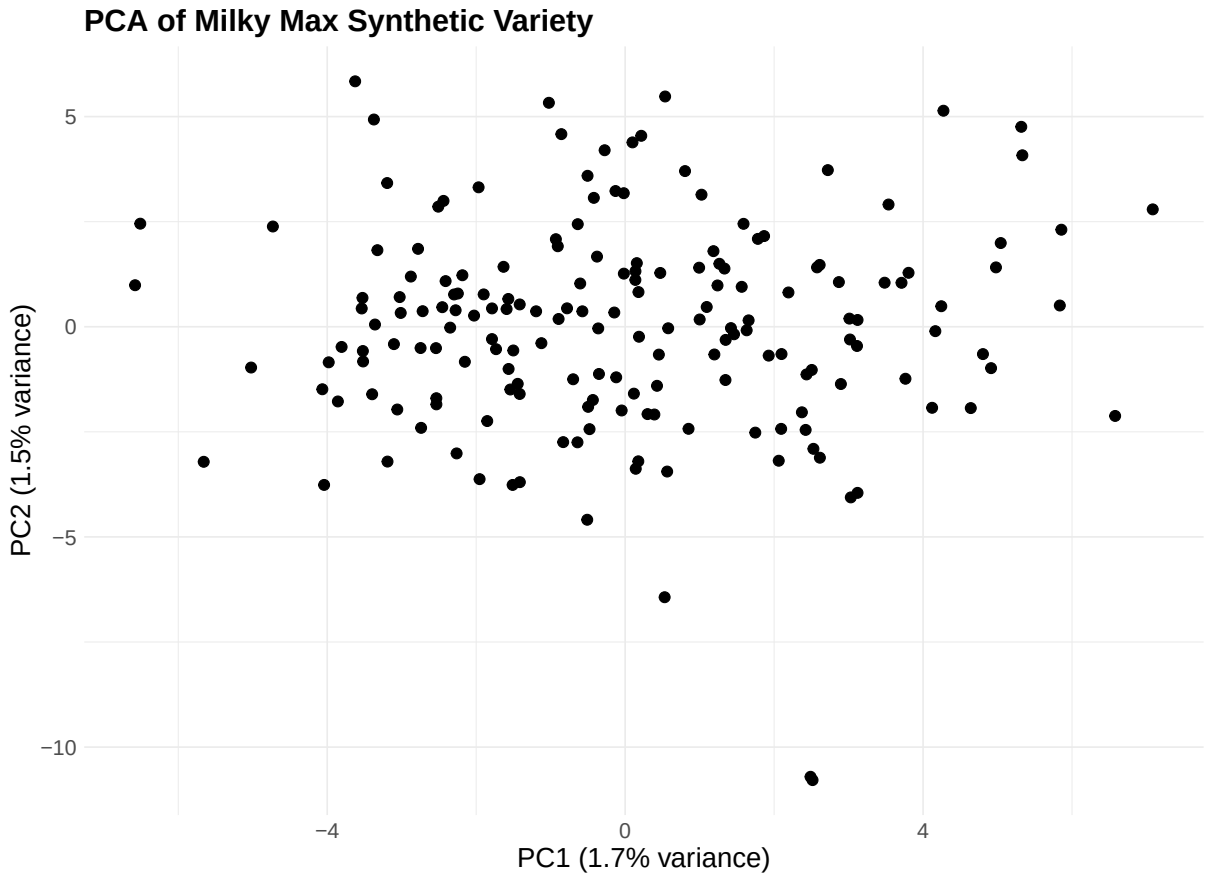
Variable	Variety (P-value)	Block (P-value)
PNpI	<0.001	<0.001
IW	<0.001	0.0126
SNpI	<0.001	0.671
SWpI	<0.001	0.395
SWpP	<0.001	<0.001
SNpP	<0.001	<0.001
TSW	<0.001	0.00151

5.3 Genetic Structure

Appendix 5: PCA for P27 populations.



Appendix 6: PCA for Milky Max.



5.4 Variances and Heritabilities

Appendix 7: Variance Components and Heritability of Seed Production Traits in Alfalfa for F1-2023 population.

Trait	Spatial	Genetic	Residual	Heritability
Pod num./Inflo.	1.88×10^{0}	5.61×10^{0}	1.32×10^{1}	0.27
Inflo. weight	2.23×10^{-4}	3.34×10^{-3}	3.26×10^{-3}	0.49
Seed num./inflo.	2.07×10^{1}	1.81×10^{2}	1.18×10^{2}	0.56
Seed weight/inflo.	5.12×10^{-5}	7.35×10^{-4}	5.09×10^{-4}	0.56
Seed weight/pod	9.52×10^{-7}	1.49×10^{-6}	4.70×10^{-7}	0.51
Seed number/pod	3.56×10^{-1}	3.24×10^{-1}	3.70×10^{-2}	0.45
Thousand seed weight	6.38×10^{-3}	1.83×10^{-2}	5.29×10^{-2}	0.23

Appendix 8: Variance Components and Heritability for Traits Related to Seed Production in Alfalfa for MilkyMax-2023 population.

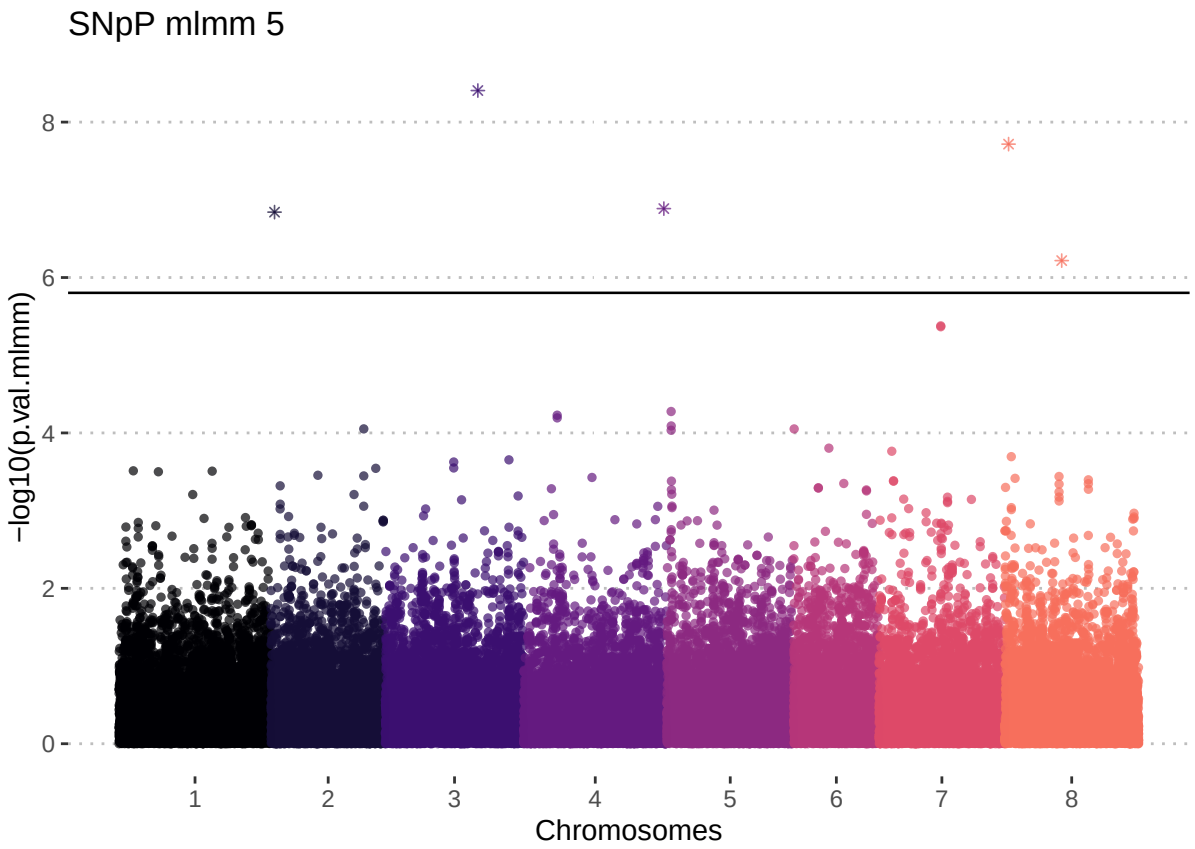
Trait	**Spatial**	**Genetic**	**Residual**	**Heritability**
Pod num./Inflo.	7.19×10^{0}	7.34×10^{0}	1.19×10^{1}	0.28
Inflo. weight	3.42×10^{-3}	1.39×10^{-3}	3.75×10^{-3}	0.16
Seed num./inflo.	8.96×10^{1}	1.17×10^{2}	1.25×10^{2}	0.35
Seed weight/inflo.	1.17×10^{-3}	5.21×10^{-5}	3.06×10^{-5}	0.04
Seed weight/pod	4.43×10^{-7}	6.44×10^{-7}	1.80×10^{-6}	0.22
Seed number/pod	1.84×10^{-1}	3.65×10^{-1}	1.94×10^{-1}	0.49
Thousand seed weight	1.87×10^{-2}	1.08×10^{-1}	1.78×10^{-3}	0.84

Appendix 9: Genetic Variances P27-2023 After Spatial Correction

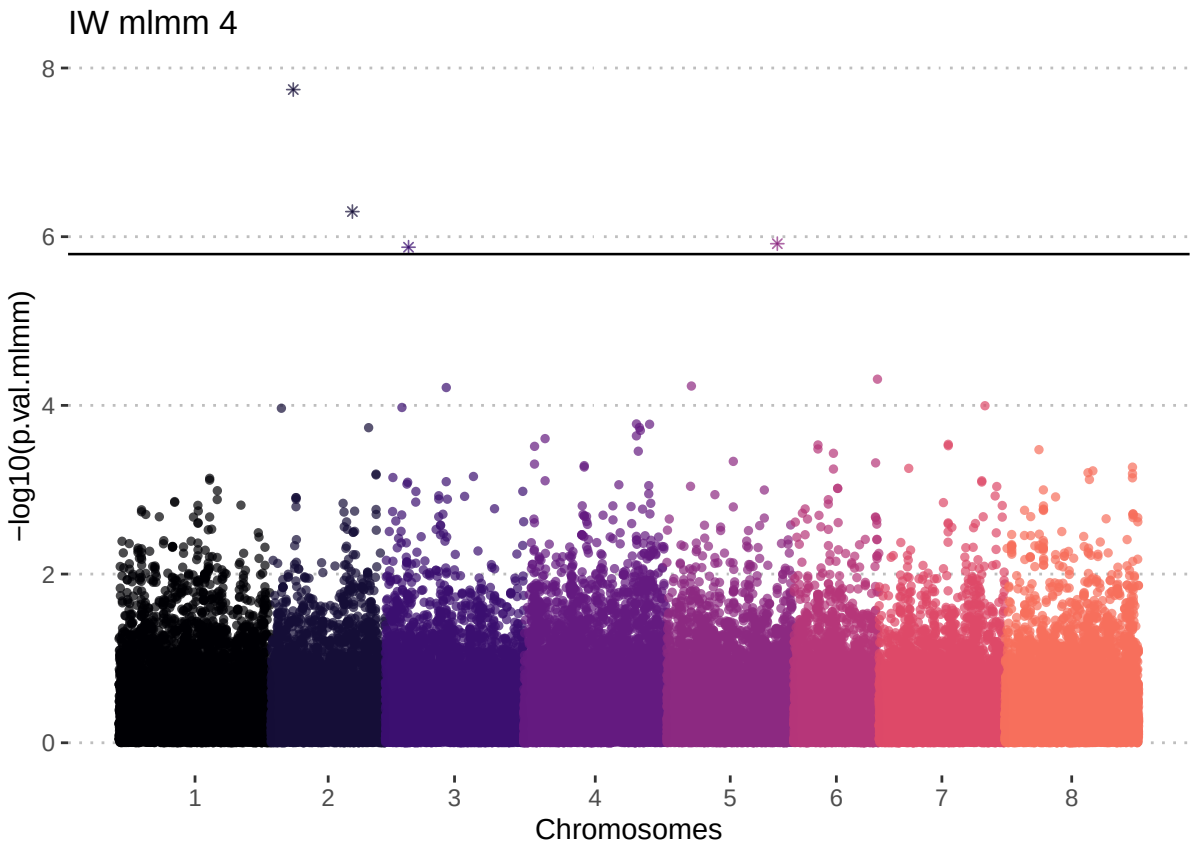
Trait	**Genetic**	**Spatial**	**Residual**	**Heritability**
Pod num./Inflo.	6.993×10^{0}	1.085×10^{-2}	8.246×10^{0}	0.46
Inflo. weight	2.235×10^{-3}	3.000×10^{-6}	2.244×10^{-3}	0.50
Seed num./inflo.	75.930×10^{0}	0.1494×10^{0}	107.600×10^{0}	0.41
Seed weight/inflo.	2.870×10^{-4}	1.000×10^{-6}	4.420×10^{-4}	0.39
Seed weight/pod	2.422×10^{-2}	4.910×10^{-3}	0.1229×10^{0}	0.16
Seed number/pod	0.1851×10^{0}	2.370×10^{-4}	0.3946×10^{0}	0.32
Thousand seed weight	3.493×10^{-2}	9.800×10^{-5}	6.862×10^{-2}	0.34

5.5 GWAS

Appendix 10: Manhattan plot for MilkyMax-2023 population for Seed Number per Pod trait.



Appendix 11: Manhattan plot for F1-2023 population for Inflorescence Weight trait.



Appendix 12: Summary of Genomic Associations Identified for Various Traits Across Different Datasets. The table lists significant markers, their associated traits, r2 values, and the dataset in which they were identified.

Marker	Trait	r2	Dataset
chr6.2_452788	adj_SNpI	0.025	P27-22/23
chr1.2_36542070	adj_SWpI	0.023	P27-22/23
chr1.2_36542070	adj_SWpP	0.028	P27-22/23
chr4.2_73304915	adj_SNpP	0.028	P27-22/23
chr6.2_84468503	adj_TSW	0.022	P27-22/23
chr3.2_79337878	adj_IW	0.031	P27-23
chr1.2_52718807	adj_SNpI	0.035	P27-23
chr1.2_52718807	adj_SWpI	0.036	P27-23
chr1.2_52718807	adj_SWpP	0.049	P27-23
chr1.2_52718807	adj_SNpP	0.045	P27-23
chr2.2_21391439	IW	0.208	F1-23WA
chr5.2_77029282	IW	0.158	F1-23WA
chr3.2_28840119	IW	0.159	F1-23WA
chr2.2_60522359	IW	0.132	F1-23WA
all	IW	0.427	F1-23WA
chr8.2_72610801	TSW	0.121	F1-23WA
chr8.2_11767356	SNpP	0.186	MilkyMax-23WA
chr3.2_71781907	SNpP	0.208	MilkyMax-23WA
chr2.2_11993317	SNpP	0.053	MilkyMax-23WA
chr4.2_92760839	SNpP	0.069	MilkyMax-23WA
chr8.2_36886928	SNpP	0.147	MilkyMax-23WA
all	SNpP	0.47	MilkyMax-23WA

Glossary

Akaike Information Criterion (AIC) is a method used to compare different statistical models. It is particularly useful in models where parameters (like the number of knots in B-splines) need to be chosen optimally. AIC provides a measure of the quality of each model, balancing the goodness of fit with the simplicity of the model. The model with the lowest AIC is generally preferred because it suggests the best trade-off between fitting the data well and keeping the model sufficiently simple. 11

B-splines are a series of polynomials used in numerical and statistical modeling to create smooth curves that can adapt flexibly to different shapes of data. The use of B-splines in the context of field trials involves creating a smooth, continuous surface to represent how traits vary spatially within a plot. A tensor product of B-splines extends this idea to two dimensions (e.g., rows and columns of a field). Here, separate B-spline bases are applied to each dimension, and their tensor product creates a grid-like structure of spline surfaces. This method allows for capturing more complex spatial patterns that might occur due to variability in soil quality, moisture, sunlight exposure, and other environmental factors across a plot. 11

Random Knot Effects (RKE). Knots in a spline are points that help in defining the piecewise polynomial segments of the spline. Typically, the placement of these knots can significantly influence the shape and flexibility of the resulting spline surface. In statistical modeling, particularly with RKE, the knots' positions are treated as random variables. This means they are not fixed in advance but are instead estimated from the data. A tensor product of B-splines extends this idea to two dimensions (e.g., rows and columns of a field). Here, separate B-spline bases are applied to each dimension, and their tensor product creates a grid-like structure of spline surfaces. This method allows for capturing more complex spatial patterns that might occur due to variability in soil quality, moisture, sunlight exposure, and other environmental factors across a plot. 11

Restricted Maximum Likelihood (REML) is a statistical method used to estimate the parameters of a linear mixed-effects model, like the one presented in Equation 3. By focusing on estimating the variance components rather than the mean structure, REML achieves more accurate and unbiased estimates of these components. This is particularly crucial in analyzing data where variability due to random factors, such as genetic differences or specific environmental effects in field trials, is of interest. In mixed models, REML's estimation of random effects' variance components forms the basis for making reliable predictions and inferences, helping to understand the underlying patterns in the data effectively. 11

species complex is defined as a group of related taxa with the close morphology to a degree that demarcation is not very clear. The term in it is broadest sense includes various concepts such as cryptic species, sibling species, species flock, superspecies, species aggregate, and *sensu lato*. 2

taxonomic continuum refers to a perspective in biological taxonomy where the boundaries between different taxonomic categories (like species, genera, or families) are seen as fluid or gradational rather than fixed or discrete. This concept acknowledges that the categorization of organisms into distinct groups, which traditionally depends on certain measurable or observable characteristics, can sometimes be arbitrary or insufficient due to the gradual nature of evolutionary changes. A taxonomic continuum highlights the idea that biological diversity is continuous, with transitional forms often existing between recognized groups. This can make it challenging to strictly define where one species ends and another begins, especially in cases where there is significant hybridization, gene flow, or in groups where evolutionary changes are particularly gradual. 2

References

Annicchiarico, P., Nazzicari, N., Li, X., Wei, Y., Pecetti, L., & Brummer, E. C. (2015). Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC Genomics*, *16*(1), 1020. https://doi.org/10.1186/s12864-015-2212-y

Bolaños-Aguilar, E. (2001, September 24). *Etude génétique de la production de graines chez la luzerne (Medicago sativa L.)* [Doctoral dissertation, École Nationale Supérieure Agronomique de Rennes].

Bolaños-Aguilar, E., Huyghe, C., Julier, B., & Ecalle, C. (2000). Genetic variation for seed yield and its components in alfalfa (medicago sativa l.) populations. *Agronomie*, *20*(3), 333–345. https://doi.org/10.1051/agro:2000131

Bouton, J. H., & Gates, R. N. (2003). Grazing-tolerant alfalfa cultivars perform well under rotational stocking and hay management. *Agronomy Journal*, *95*(6), 1461–1464. https://doi.org/10.2134/agronj2003.1461

Carof, M. (2006, October). *Fonctionnement de peuplements en semis direct associant du blé tendre d'hiver (Triticum aestivum L.) à différentes plantes de couverture en climat tempéré* (Publication No. 2006INAP0016) [Theses]. INAPG (AgroParisTech). https://pastel.hal.science/tel-00338817

Carrere, S., & Gouzy, J. (2024). Xinjiang-daye alfalfa variety genome browser [Accessed: 2024-06]. https://bbric-pipelines.toulouse.inra.fr/myGenomeBrowser?portalname=MSAT_XinJiangDaYe&owner=sebastien.carrere@inrae.fr&key=PyG9k9tK

Cericola, F., Lenk, I., Fè, D., Byrne, S., Jensen, C. S., Pedersen, M. G., Asp, T., Jensen, J., & Janss, L. (2018). Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (lolium perenne l.) *Frontiers in Plant Science*, *9*, 369. https://doi.org/10.3389/fpls.2018.00369

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F., Liu, W., Li, Y., Liu, Y., Zhang, X., Zhang, R., Zhang, Y., Li, Y., Wang, K., He, H., Wang, Z., Fan, G., Yang, H., ... Qiu, Q. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature Communications*, *11*(1), 2494. https://doi.org/10.1038/s41467-020-16338-x

Cougnon, M., Durand, J., Julier, B., et al. (2022). Using perennial plant varieties for use as living mulch for winter cereals. a review. *Agronomy for Sustainable Development*, *42*, 110. https://doi.org/10.1007/s13593-022-00844-x

Deb, D. (2021). Productive efficiency of traditional multiple cropping systems compared to monocultures. *Experimental Results*. https://doi.org/doi.org/10.1017/exp.2021.7

Demirdogen, A., Guldal, H. T., & Sanli, H. (2023). Monoculture, crop rotation policy, and fire. *Ecological Economics*, *203*, 107611. https://doi.org/10.1016/j.ecolecon.2022.107611

Dong, N., Sun, Y., Guo, T., et al. (2020). Udp-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nature Communications*, *11*, 2629. https://doi.org/10.1038/s41467-020-16403-5

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species (L. Orban, Ed.). *PLoS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Fernandez, A., Sheaffer, C., Tautges, N., Putnam, D., & Hunter, M. (2019). Alfalfa, wildlife & the environment.

Flajoulot, S., Ronfort, J., Baudouin, P., Barre, P., Huguet, T., Huyghe, C., & Julier, B. (2005). Genetic diversity among alfalfa (medicago sativa) cultivars coming from a breeding program, using SSR markers. *Theoretical and Applied Genetics*, *111*(7), 1420–1429. https://doi.org/10.1007/s00122-005-0074-4

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, *54*(1), 357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Gallais, A. (2003). *Quantitative genetics and breeding methods in autopolyploid plants* (1st ed.). Quae.

Gardner, K. M., & Latta, R. G. (2007). Shared quantitative trait loci underlying the genetic correlation between continuous traits. *Molecular Ecology*, *16*(20), 4195–4209. https://doi.org/10.1111/j.1365-294X.2007.03499.x

Gharabli, H., Della Gala, V., & Welner, D. H. (2023). The function of udp-glycosyltransferases in plants and their possible use in crop protection. *Biotechnology Advances*, *67*. https://doi.org/https://doi.org/10.1016/j.biotechadv.2023.108182

Hanson, A. A., Barnes, D. K., & Hill Jr., R. R. (Eds.). (1988). *Alfalfa and alfalfa improvement*. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc. https://doi.org/10.2134/agronmonogr29

He, F., Zhang, F., Jiang, X., Long, R., Wang, Z., Chen, Y., Li, M., Gao, T., Yang, T., Wang, C., Kang, J., Chen, L., & Yang, Q. (2022). A genome-wide association study coupled with a transcriptomic analysis reveals the genetic loci and candidate genes governing the flowering time in alfalfa (medicago sativa l.) *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.913947

Herrmann, D., Barre, P., Santoni, S., et al. (2010). Association of a CONSTANS-LIKE gene to flowering and height in autotetraploid alfalfa. *Theoretical and Applied Genetics*, *121*, 865–876. https://doi.org/10.1007/s00122-010-1356-z

Huang, X., Liu, L., Qiang, X., Meng, Y., Li, Z., & Huang, F. (2024). Integrated multi-omics analysis to reveal the molecular mechanisms of inflorescence elongation in medicago

sativa. *International Journal of Molecular Sciences*, *25*(12), 6497. https://doi.org/10.3390/ijms25126497

Huss, C., Holmes, K., & C, B. (2022). Benefits and risks of intercropping for crop resilience and pest management. *Journal of Economic Entomology*. https://doi.org/doi.org/10.1093/jee/toac045

Jefferson, P. G., & Cutforth, H. W. (2005). Comparative forage yield, water use, and water use efficiency of alfalfa, crested wheatgrass and spring wheat in a semiarid climate in southern saskatchewan. *Canadian Journal of Plant Science*, *85*(4), 877–888. https://doi.org/10.4141/P04-115

Julier, B., Gastal, F., Louarn, G., Badenhausser, I., Annicchiarico, P., Crocq, G., Chatelier, D. L., Guillemot, E., & Emile, J. C. (2017, January). Lucerne (alfalfa) in european cropping systems. In D. Murphy-Bokern, F. L. Stoddard, & C. A. Watson (Eds.), *Legumes in cropping systems* (1st ed., pp. 168–192). CABI. https://doi.org/10.1079/9781780644981.0168

Julier, B., Barre, P., Lambroni, P., Delaunay, S., Thomasset, M., Lafaillette, F., & Gensollen, V. (2018). Use of GBS markers to distinguish among lucerne varieties, with comparison to morphological traits. *Molecular Breeding*, *38*(11), 133. https://doi.org/10.1007/s11032-018-0891-1

Julier, B., Huyghe, C., & Ecalle, C. (2000). Within and among-cultivar genetic variation in alfalfa: Forage quality, morphology, and yield. *Crop Science - CROP SCI*, *40*. https://doi.org/10.2135/cropsci2000.402365x

Karancsi, Z. (2010). Agriculture: Deforestation. In J. Szabó, L. Dávid, & D. Lóczy (Eds.), *Anthropogenic geomorphology* (pp. 95–112). Springer Netherlands. https://doi.org/10.1007/978-90-481-3058-0_8

Labreuche, J., Hauprich, P., Bodilis, A., & Soenen, B. (2017). Impact de couverts permanents sur le ble tendre d'hiver.

Lawrence-Paul, E. H., Poethig, R. S., & Lasky, J. R. (2023). Vegetative phase change causes age-dependent changes in phenotypic plasticity. *New Phytologist*, *240*(2), 613–625. https://doi.org/https://doi.org/10.1111/nph.19174

Lê, S., Josse, J., & Husson, F. (2008). Factominer: An r package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Lefebvre, H. (2023, June). *Déterminisme génétique de la production de semences chez la luzerne, plante fourragère* [Master's thesis, URP3F, INRAE Centre de recherche Nouvelle Aquitaine-Poitiers, Le Chêne, RD 150, CS 80006, 86600 Lusignan]. https://dumas.ccsd.cnrs.fr/dumas-04357944

Lesins, K. A., & Lesins, I. (1979). *Genus medicago (leguminosae)*. Springer Netherlands. https://doi.org/10.1007/978-94-009-9634-2

Li, X., & Brummer, E. C. (2012). Applied genetics and genomics in alfalfa breeding. *Agronomy*, *2*(1), 40–61. https://doi.org/10.3390/agronomy2010040

Lin, C.-Y., Xing, G., & Xing, C. (2012). Measuring linkage disequilibrium by the partial correlation coefficient. *Heredity*, *109*(6), 401–402. https://doi.org/10.1038/hdy.2012.54

Long, R., Zhang, F., Zhang, Z., Li, M., Chen, L., Wang, X., Liu, W., Zhang, T., Yu, L.-X., He, F., Jiang, X., Yang, X., Yang, C., Wang, Z., Kang, J., & Yang, Q. (2022). Assembly of chromosome-scale and allele-aware autotetraploid genome of the chinese alfalfa cultivar

zhongmu-4 and identification of snp loci associated with 27 agronomic traits. *bioRxiv*, 428692. https://doi.org/10.1101/2021.02.21.428692

Marcos, Zornoza, & Sánchez. (2020). Asociacion de cultivos en horticultura para incrementar la productividad agraria y los servicios ecosistémicos. https://doi.org/10.31428/10317/10300

Medina, C. A., Hawkins, C., Liu, X.-P., Peel, M., & Yu, L.-X. (2020). Genome-wide association and prediction of traits related to salt tolerance in autotetraploid alfalfa (medicago sativa l.) *International Journal of Molecular Sciences*, *21*(9), 3361. https://doi.org/10.3390/ijms21093361

Mei, L., Zhang, N., Wei, Q., Cao, Y., Li, D., & Cui, G. (2022). Alfalfa modified the effects of degraded black soil cultivated land on the soil microbial community. *Frontiers in Plant Science*, *13*, 938187. https://doi.org/10.3389/fpls.2022.938187

Meiss, H., Médiène, S., Waldhardt, R., Caneill, J., & Munier-Jolain, N. (2010). Contrasting weed species composition in perennial alfalfas and six annual crops: Implications for integrated weed management. *Agronomy for Sustainable Development*, *30*(3), 657–666. https://doi.org/10.1051/agro/2009043

MNHN & OFB [Ed]. (2024). Fiche de Medicago sativa L., 1753 [Accessed: June 13, 2024]. *Inventaire national du patrimoine naturel (INPN)*.

Mu, L., Su, K., Zhou, T., & Yang, H. (2023). Yield performance, land and water use, economic profit of irrigated spring wheat/alfalfa intercropping in the inland arid area of northwestern china. *Field Crops Research*, *303*, 109116. https://doi.org/10.1016/j.fcr.2023.109116

Muller, M. H., Prosperi, J. M., Santoni, S., & Ronfort, J. (2003). Inferences from mitochondrial DNA patterns on the domestication history of alfalfa (*Medicago sativa*). *Molecular Ecology*, *12*(8), 2187–2199. https://doi.org/10.1046/j.1365-294X.2003.01897.x

Muñoz, F., & Sánchez, L. (2024). *Breedr: Statistical methods for forest genetic resources analysts* [R package version 0.12-5]. https://github.com/famuvie/breedR

Osterholz, W., Renz, M., Jokela, W., & Grabber, J. (2019). Interseeded alfalfa reduces soil and nutrient runoff losses during and after corn silage production. *Journal of Soil and Water Conservation*, *74*(1), 85–90. https://doi.org/10.2489/jswc.74.1.85

Pégard, M., Barre, P., Delaunay, S., Surault, F., Karagić, D., Milić, D., Zorić, M., Ruttink, T., & Julier, B. (2023). Genome-wide genotyping data renew knowledge on genetic diversity of a worldwide alfalfa collection and give insights on genetic control of phenology traits. *Frontiers in Plant Science*, *14*, 1196134. https://doi.org/10.3389/fpls.2023.1196134

Petit, C., & Aubry, C. (2012, June). Géographie de l'agriculture biologique et opportunités locales de débouchés commerciaux pour les exploitations agricoles [Poster]. *Institut Technique de l'Agriculture Biologique (ITAB). PARIS, FRA*. https://hal.science/hal-01190513

Prosperi, J.-M., Jenczewski, E., Muller, M.-H., Fourtier, S., Sampoux, J.-P., & Ronfort, J. (2014). Alfalfa domestication history, genetic diversity and genetic resources. (4).

Qiang, H., Chen, Z., Zhang, Z., Wang, X., Gao, H., & Wang, Z. (2015). Molecular diversity and population structure of a worldwide collection of cultivated tetraploid alfalfa (medicago sativa subsp. sativa l.) germplasm as revealed by microsatellite markers (Z. Wang, Ed.). *PLOS ONE*, *10*(4), e0124592. https://doi.org/10.1371/journal.pone.0124592

Quiros, C. F., & Bauchan, G. R. (1988). The genus medicago and the origin of the medicago sativa comp. In A. A. Hanson, D. K. Barnes, & R. R. Hill (Eds.), *Agronomy monographs* (pp. 93–124). American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. https://doi.org/10.2134/agronmonogr29.c3

Şakiroğlu, M., & İlhan, D. (2021). *Medicago sativa* species complex: Revisiting the century-old problem in the light of molecular tools. *Crop Science*, *61*(2), 827–838. https://doi.org/10.1002/csc2.20316

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, *44*(7), 825–830. https://doi.org/10.1038/ng.2314

Shen, C., Du, H., Chen, Z., Lu, H., Fugui, Z., Chen, H., Meng, X., Liu, Q., Liu, P., Zheng, L., Li, X., Dong, J., Liang, C., & Tao, W. (2020). The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasms provide genomic resources for alfalfa research. *Molecular Plant*, *13*. https://doi.org/10.1016/j.molp.2020.07.003

Sul, J. H., Martin, L. S., & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genetics*, *14*(12), e1007309. https://doi.org/10.1371/journal.pgen.1007309

Teuber, L. R., & Brick, M. A. (1988). Morphology and anatomy. In *Alfalfa and alfalfa improvement* (pp. 125–162). John Wiley & Sons, Ltd. https://doi.org/10.2134/agronmonogr29.c4

VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, *9*, 255–266. https://doi.org/10.1038/nrg2322

Wall, K. G., & Steppuhn, H. (2007). Does seed size affect alfalfa establishment and productivity in saline seedbeds? *University of Saskatchewan*. https://harvest.usask.ca/server/api/core/bitstreams/607ac0bc-5b57-409a-8dcd-92ab6bd0a1bd/content