ವಿಶ್ವೇಶ್ವರಯ್ಯ ತಾಂತ್ರಿಕ ವಿಶ್ವವಿದ್ಯಾಲಯ, ಬೆಳಗಾವಿ
VISVESVARAYA TECHNOLOGICAL UNIVERSITY - BELAGAVI

Ph. D Dissertation on

"DEVELOPMENT OF DISTINCT SNP GENOTYPIC ARRAYS FOR ASSOCIATION MAPPING OF GERMPLASM AND BI-PARENTAL POPULATION IN MULBERRY"

Submitted by

SINSHA P S
USN-6RV18MBT01

RESEARCH CENTRE

RASHTREEYA SIKSHANA SAMITHI TRUST ® RV INSTITUTIONS

Department of Biotechnology

R.V. College of Engineering, Bengaluru-560059

Under the Guidance of

Dr. VIDYA NIRANJAN

Professor and Head of Department- Biotechnology

R.V. College of Engineering

DECLARATION

I, **SINSHA P S** bearing the USN 6RV18MBT01, hereby declare that the research work entitled "**DEVELOPMENT OF DISTINCT SNP GENOTYPIC ARRAYS FOR ASSOCIATION MAPPING OF GERMPLASM AND BI-PARENTAL POPULATION IN MULBERRY**" carried out by me has not previously formed the basis for the award of any degree or diploma or certificate from any other university. The thesis is based on the individual original research work. The thesis does not contain any material that infringes the copyright of any individual or organization. The text, tables, equations, figures, charts and graphs taken from sources such as research articles, books, periodicals, websites have been cited appropriately.

Signature of the Research Scholar

SINSHA P S
USN: 6RV18MBT01
Department of Biotechnology
RVCE, Bengaluru – 560059.

RV COLLEGE OF ENGINEERING

RV Vidyaniketen Post, 8th Mile, Mysuru Road

Bengaluru – 560059

CERTIFICATE

This is to certify that the thesis entitled **"Development of Distinct SNP Genotypic Arrays for Association mapping of Germplasm and Bi-parental Population in Mulberry"** is carried out by me, **Sinsha P S, USN : 6RV18MBT01** a bonafide student of R V College of Engineering, in partial fulfilment for the award of the degree of Doctor of Philosophy in the Faculty of Bioinformatics of Visvesvaraya Technological University, Belgavi during the year 2024.

To the best of my knowledge, the work reported in this thesis has not been submitted by me elsewhere for the award of any degree and is not a repetition of work carried out by others.

Sinsha

Sinsha P S

Research Scholar

Vidya Niranjan

12/3/24

Head of the Department
Dept. of Bio-Technology
R.V. College of Engineering
BANGALORE - 560 059

Signature of the Guide

Dr. Vidya Niranjan

Professor & HOD

Department of Biotechnology

R V College of Engineering

Bengaluru- 560059

Signature of the Principal

Dr. K. N. Subramanya

R V College of Engineering

Bengaluru- 560059

PRINCIPAL
RV COLLEGE OF ENGINEERING
BENGALURU - 560 059

ACKNOWLEDGEMENTS

With profound appreciation and reverence, I would like to extend my heartfelt gratitude to my Ph.D. project advisor, **Dr. Vidya Niranjan**, Professor and Head of the Department of Biotechnology and Centre of Excellence in Computational Genomics. Her invaluable assistance, unwavering encouragement, and extensive expertise have been instrumental in guiding me throughout the entirety of my doctoral journey.

I would also like to express my sincere acknowledgment to **Dr. K N Subramanya**, Principal of R V College of Engineering, for fostering an environment conducive to the successful execution of my research project. Furthermore, I am grateful to the teaching and non-teaching staff of the Department of Biotechnology for their unwavering assistance and support in the laboratory, which proved vital in the completion of my research work.

I would like to express my sincere gratitude to my lab mates at **Centre of Excellence Computational Genomics** for their unwavering support and assistance throughout the course of my doctoral research. Their collaboration, encouragement, and camaraderie have been invaluable to me during this journey.
I am deeply thankful to **Lavanya C, Spoorthi R Kulkarni, Shreya Satyanarayan Bhat, Akshay Uttarkar C, Anagha S Setlur, Chandrashekar K, S Pooja** for their willingness to share their expertise, provide constructive feedback, and engage in stimulating discussions. Their diverse perspectives and dedication to excellence have enriched my research experience immeasurably.
I am also grateful for the countless hours spent together in the lab, troubleshooting experiments, analyzing data, and brainstorming ideas. The sense of camaraderie and teamwork fostered within our lab has been instrumental in overcoming challenges and achieving milestones.

I am immensely grateful to the **Central Sericulture Research and Training Institute, Central Silk-Board, Mysore**, for their invaluable assistance. I would like to extend my sincere appreciation to **Dr. V. Sivaprasad**, the project coordinator, as well as the other esteemed investigators, **Dr. S. Manthira Moorthy, Dr. K.P Arunkumar, Dr. Kusuma L, and Dr. R. Sumathy**, for generously providing us with silkworms for RAD sequencing. Their support and contribution have been instrumental in the success of our research, and I am deeply thankful for their assistance.

We express our sincere gratitude to the **Department of Biotechnology** for their invaluable support in facilitating the project titled "**Genetic Enhancement of Mulberry by Genomics Approach: A Multi Component Network Project**." Their unwavering commitment to advancing scientific research has been instrumental in driving our exploration towards innovative solutions in agricultural genomics.

Finally, I would like to express my gratitude to my family and friends for their unwavering moral support during the research work.

SINSHA P S

6RV18MBT01

ABSTRACT

This research presents a comprehensive analysis of four Mulberry genotypes, each harboring unique traits crucial for sericulture and broader agricultural applications. Through meticulous investigation, genetic variations, including single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs), were uncovered in the Thailand Male genotype, renowned for its high-quality leaves. These findings offer insights into the genetic factors associated with superior leaf yield, laying the groundwork for selective breeding to develop Mulberry cultivars with consistently nutrient-rich leaves for silkworms. Additionally, the study elucidated the genetic basis for the resistance to root rot diseases in the Assam Bola genotype, originating from India, potentially facilitating the development of disease-resistant Mulberry cultivars, thereby promoting environmentally sustainable practices. The research examined the drought-resistant S1 genotype, aiming to understand its exceptional resilience under water scarcity conditions, critical for ensuring a stable leaf supply amidst increasing drought occurrences due to climate change. Additionally, a comprehensive SNP and SSR database was generated, serving as a valuable resource for future genetic studies in Mulberry. The study also investigated effective nitrogen utilization in Mulberry, utilizing the Punjab Local genotype, a vital aspect in enhancing agricultural productivity while minimizing environmental impact. Utilizing state-of-the-art methods, such as SNP and SSR markers, a vast array of genetic data was generated, encompassing approximately 50 million SNPs. To streamline the data, consensus generation was conducted, facilitating the development of diagnostic tools and genetic linkage maps. Moreover, primers were designed for chip development, enabling further genetic analysis and enhancement of Mulberry cultivars. Phenotype-genotype correlation studies were successfully achieved, providing valuable insights into the genetic mechanisms underlying key agronomic traits. Additionally, phenotype-phenotype correlation studies further enhanced our understanding of the complex interactions between genetic variations and phenotypic traits in Mulberry. Overall, this research contributes to the advancement of Mulberry agriculture by identifying genetic markers, elucidating genetic linkages, and creating diagnostic tools, ultimately enhancing economic viability and environmental responsibility in the sericulture industry.

Key words: Morus indica, Markers, Single Nucleotide Polymorphism, Simple sequence repeats, Genotype, Phenotype, Genetic mapping

CHAPTER-WISE EXPLANATION

This thesis is presented in nine different chapters which are as follows:

Chapter-1 Introduction

Introduction summarizes the importance of Mulberry, Specific details of *Morus indica*, Economic importance, Medicinal value and the Phenotypic details of *Morus indica*

Chapter-2 Review of Literature

Presents the previous research on related topics, the literature gap, steps that have current study, and motivation and scope of study.

Objectives

Presents the frame research objectives of the study.

Chapter-3 Research Methodology

Includes all the materials and methods that were used in the study to arrive at the expected outcomes.

Chapter -4 Results and Discussion

Presents comprehensive results that were obtained from each methodology, including some value-added results. The results include the quality checks, adaptor trimming, alignment, variant calling, analysis of mutational profiles, data normalization, analysis of unique variants, database development and design of mulberry array with markers. Discussions for each result by comparing with previous results are also presented.

Chapter 5- Conclusions and future perspectives

Summarizes the findings of this research work and outlines specific conclusions drawn from the experimental efforts. Further, the future scope of the present work is provided.

Chapter 6- References

All the reference has been cited

LIST OF PUBLICATIONS

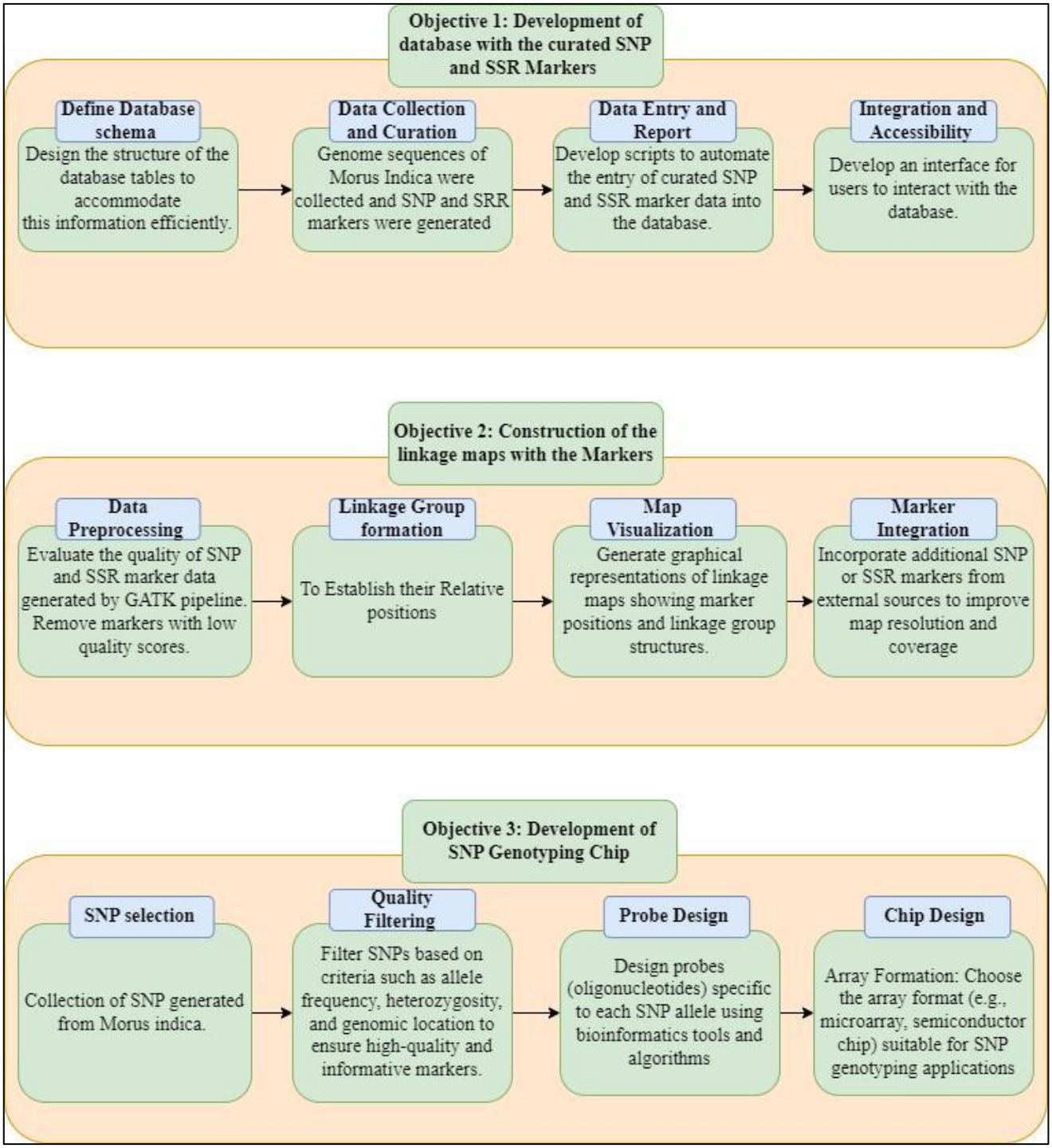
1) Sinsha Vikhin, Lavanya C, Shri Ganapathi V Raman, Anirudh R URS, Shashank Rao Vidya Niranjan "COMPUTATIONAL TOOLS FOR UNVEILING MULBERRY'S GENETIC ARCHITECTURE THROUGH LINKAGE MAPPING: A COMPREHENSIVE REVIEW", Gradiva Review Journal, ISSN NO: 0363-8057, Volume 9, ISSUE 8, 2023, DOI: 10.37897.GRJ. 2022.V9I8.23.513311

2) Sinsha Vikhin, Vidya Niranjan , Lavanya C, Spoorthi R Kulkarni, "Unlocking the Genetic Diversity of *Morus indica* through Molecular Marker Analysis", JOURNAL OF TECHNOLOGY, ISSN NO: 1012-3407, VOL 13, ISSUE 10, DOI: 10.61350/v13-105416

LIST OF FIGURES

LIST OF TABLES

OVERVIEW OF THE STUDY



CHAPTER 1

1.0 INTRODUCTION

The mulberry, belonging to the *Moraceae* genus, is a fast-growing, woody, hardy perennial plant. It has several crucial qualities, including a larger leaf output, a shorter gestation time, and better environmental tolerance. More than 15 species of the genus are mostly found in temperate, tropical, and subtropical climates in Asia, Africa, and North America. The following species: *Morus alba, nigra, rubra, and australis* More than 15 species of the genus are mostly found in temperate, tropical, and subtropical climates in Asia, Africa, and North America. Among these, *Morus alba, Morus nigra, Morus rubra, and Morus australis, Morus atropurpurea, Morus cathayana,* and *Morus notabilis* are some of the more notable ones [1].

1.1. TAXONOMY AND MAJOR SPECIES

The fruits, which mature to white, pink, red, or purple, resemble blackberries in several ways. Individuals can be dioecious or monoecious, having both male and female flowers (bearing only male or female flowers) One metric ton of mulberry leaves is thought to be required for the raising of silkworms that emerge from one case of eggs and produce roughly 25 kg to 30 kg of high-quality cocoons. Various species of mulberry, including *Morus indica* and *Morus mesosygia*, among others, are renowned for their economic importance. Mulberry leaves are commonly utilized to nourish *Bombyx mori* silkworms, essential for silk production. Through a process wherein the silkworm converts mulberry leaf protein into sericin and fibroin, the foundation of silk is established. This silk is then commercially utilized in the creation of fine clothing. Globally, mulberry silk constitutes approximately 90% of all raw silk production, significantly benefiting countless individuals [2]. Additionally, the mulberry plant offers various nutritional and therapeutic advantages. Many nations have historically employed mulberry leaves and fruits, abundant in protein and vitamins, as both animal feed and human sustenance. Mulberry cultivation primarily supports silkworm production, serving as the exclusive food source during their larval stage. Therefore, maintaining a robust production of mulberry leaves according to scientific standards is crucial for sustainable sericulture. Mulberry trees are deciduous, featuring alternately positioned, toothed, and occasionally lobed leaves. Clusters of small blooms, called catkins, adorn the branches, each developing into a fruit cluster known as a "multiple." These fruits, ranging in color from white to purple, bear resemblance to blackberries. Mulberry trees can exhibit dioecious or monoecious characteristics, with some possessing separate male and female flowers while others have both on the same plant. It's estimated that one metric ton of mulberry leaves is necessary to rear silkworms hatching from a single batch of eggs, yielding approximately 25 kg to 30 kg of premium-quality cocoons. The genus Morus has 68 species, the majority of which are found in Asia [3-5]. Over

a thousand different types are grown in China. Descended from four primary species—Guangdong Mulberry *(Morus atropurpurea*), Mountain Mulberry (*Morus bombycis*), White Mulberry (*Morus alba*), and Lu Mulberry (*Morus multicaulis*)—mulberry varieties exhibit diversity. In India, prominent species include *M. indica, M. alba, M. serrata, and M. laevigata*, predominantly found in the northern regions. Most cultivated types belong to either *M. indica or M. alba*. The Central Sericulture Research and Training Institute in Mysore, India, maintains a collection of 223 mulberry cultivars, encompassing indigenous, alien, unidentified, and top hybrid varieties. Across the former Soviet Union's Republics, M. alba and M. indica are widely recognized. Through selective breeding methods such as open pollination, controlled hybridization, selection, and mutation breeding, numerous variations—many of which are polyploid—have been developed, surpassing a thousand in some cases. In Brazil alone, there are approximately 90 distinct M. alba variants [6,7].

1.2 INTRODUCTION TO *MORUS INDICA*

Morus indica, belonging to the Morus genus within the *Moraceae* family of flowering plants, is widely recognized as part of the mulberry family. *Morus indica* is a deciduous tree. Indigenous to the temperate and subtropical Himalayan region, it is now grown in various regions including India, China, Japan, and East Africa [8] *Morus indica*, as together with more mulberry family members, is frequently characterized as either a small tree or a shrub, rarely exceeding 10–15 meters (33–49 feet). When young, the branches are covered in fine, soft hairs known as; however, the plant loses these hair as they grow older. The branches are a light gray-brown color. The leaves range from 4–12.5 centimeters (1+1⁄2–5 inches) long and 2.5–7.5 cm (1–3 in) wide, and are attached to the tree via petioles. The leaves themselves are usually ovate but sometimes lobed, coming to a narrow point, making them somewhere between caudate and acuminate. The leaves are retuse to slightly cordate, having a small lobe at the base. They are shortly serrated, with each tooth narrowing to a thin point, making them apiculate. The leaf color is dark green, with a paler underside covered in fine hairs. *Morus indica* is a monoecious flowering plant, having male and female flowers growing on the same tree, although often on distinct branches [8]. The male inflorescence is narrow, between 9–11.5 millimeters (3⁄8–7⁄16 in) long, and covered in fine hairs. The female flowers are subglobose, or just shy of spherical. They measure 6–9.5 mm (1⁄4–3⁄8 in) long. The stigma of these flowers is about 3.5 mm long with dense, short hair. The female flower, after being fertilized, forms a fleshy compound fruit known as a syncarp. This syncarp, which is black when fully ripe, looks like that of Morus nigra, commonly known as a black mulberry. occupies a pivotal position in the realm of sericulture, embodying an intricate interplay of biological, economic, ecological, and cultural significance. *Morus indica* leaves serve as the main source of sustenance for silkworms (*Bombyx mori*) throughout their life stages. These leaves provide a balanced mix of proteins, carbohydrates, vitamins, and minerals, which cater to the specific dietary requirements

of silkworm larvae [9]. This nutritional blend supports their robust growth and development and significantly impacts the silk's quality and quantity. The symbiotic partnership between *Morus indica* and silkworms has nurtured centuries of traditional wisdom, fostering a thriving sericulture industry that contributes substantially to regional economies. Beyond its economic impact, *Morus indica* catalyzes rigorous scientific research, driving innovations in cultivation techniques, leaf quality enhancement, and disease resistance. Moreover, this interdependence extends to ecological considerations, where the plant's innate disease-resistant compounds contribute to silkworm health and robustness. Cultivating *Morus indica* aligns with sustainability principles, utilizing a renewable plant resource, and carries cultural and historical weight in regions with deep sericultural roots. Ultimately, *Morus indica's* role transcends its botanical identity, weaving a tapestry of tradition, innovation, and ecological harmony that is intrinsic to the intricate fabric of sericulture.

1.3 IMPORTANCE OF FEED IN SERICULTURE

The specialized insect *Bombyx mori* predominantly consumes the rapidly growing perennial plant *Morus indica* (L.), commonly known as Indian mulberry and belonging to the *Moraceae* family. This plant holds significant importance in sericulture [10]. Feed plays a crucial role in sericulture by directly impacting the growth, development, and silk quality produced by silkworms. Mulberry leaves, the primary food source for silkworms, may result in decreased silk production if certain nutrients are lacking. The nutritional value of these leaves is influenced by various agro climatic factors, and any deficiency in nutrients hampers the silk production of silkworms. A diet rich in nutrients promotes the optimal growth and development of silkworm larvae, as well as the production of eggs. To meet the nutritional demands of numerous insects, supplementary sources of nutrients are necessary. This facilitates the evaluation of different fortification agents' significance and effects on silkworm nutrition. Recent efforts have been made to enhance both the quantity and quality of silk through various initiatives. These initiatives include nutrient supplementation, antibiotic spraying, application of juvenile hormone (JH), plant-based products, steroids, JH-mimicking techniques, and the use of plant extracts. In an endeavor to enhance both the quantity and quality of silk, mulberry leaves have been enriched with various nutrients for silkworm consumption. The practice of fortifying and supplementing mulberry leaves is a relatively recent advancement in sericulture research. The sericulture industry has strategically utilized nutritional interventions to optimize the functioning of silk glands, thereby impacting economic performance and silk production [11]. Mulberry leaves are fortified with an array of minerals, antibiotics, plant extracts, and hormones to augment their nutritional profile. As the exclusive food source for silkworm caterpillars, mulberry holds paramount importance in sericulture [12]. Sericulture represents a pivotal sector in agriculture, deeply interconnected with the cultivation of mulberry and the production of silk [13]. Factors such as intercropping with pulses, integrated pest and

disease management, and external variables like the COVID-19 pandemic collectively influence mulberry yield [14-16]. Notably, mulberry is often interplanted with pulses, particularly chickpeas.

1.4 *MORUS INDICA* HARVEST IN INDIA

Mulberry (*Morus indica* L.) is a foliage crop that is commonly cultivated in North-Eastern India with the goal of growing silkworms (*Bombyx mori* L.) in rainfed areas. [17-20]. Abiotic factors that can drastically reduce mulberry output include dryness, salt, and high temperatures. The species *Morus indica* is particularly noteworthy in the context of generating high yield and quality mulberry leaves in acid soil of North Eastern India. Mulberries are the sole crop farmed in many regions of India, and an analysis of the soil fertility there indicated acidic soils with varying levels of organic carbon and nutrient shortages [21,22]. Given that the genome of the domesticated Indian mulberry (*Morus indica* cv. K2) has been sequenced, comprehending repetitive DNA, protein-coding genes, and the study [23] highlighted the widespread nutritional inadequacies in the upper Brahmaputra area of northeastern India's soils. Agronomic traits include good propagation, early sprouting, rapid growth, high biomass output, wide adaptability to varied agro-climatic conditions, sensitivity to organic and inorganic fertilizers, and resistance to diseases and pests are important mulberry breeding needs. In other words, the most important factors are the silkworm's taste and the capacity to produce excellent cocoons. Thus, it is crucial to exploit the mulberry genetic resources that are presently available to develop mulberry varieties that are suitable for certain soil types prevalent in different agro-climatic zones. Mulberry farming has been practiced in India for a very long time, going all the way back to the beginning of time. Chinese traders and travelers contributed the talent [24,25]

1.5 NUTRITIONAL COMPOSITION ON MULBERRY *MORUS INDICA*

A single mulberry fruit typically weighs between 2.14 to 4.07 grams. Among its therapeutic components are natural and amino sugars [26]. Mulberry fruit contains various polysaccharides that significantly influence human physiology [27,28]. These fruit-derived polysaccharides have been associated with several beneficial bioactivities, including antioxidant, hypoglycemic [29,30], anti-obesity [31], anti-inflammatory, and anti-apoptotic effects [32,33]. Recent research has revealed that mulberry fruit contains nearly all essential amino acids. Spectroscopic techniques have identified six novel forms of morusimic acid, labeled A, B, C, D, E, and F which are branched-chain amino acids known to accelerate muscle growth [34]. Studies indicate that mulberry fruits contain 7.55% saturated lipids and 87.5% unsaturated fatty acids, with linoleic acid comprising the majority (79.4%) and palmitic acid following at 8.6% [35]. Turkish mulberry fruits exhibit a total lipid content of 57.3%, with palmitic acid constituting 22.4%. Mulberry fruits contain a higher proportion of polyunsaturated fatty acids compared to monounsaturated or saturated fatty acids [36]. In comparison to strawberries (72%), mulberries (76%) and ziziphus jujuba (68–72%) have a greater range of polyunsaturated fatty acids [37,38]. Linoleic acid,

a key polyunsaturated fatty acid abundant in mulberry fruit, plays a critical role in human growth, health promotion, and disease prevention [39,40].

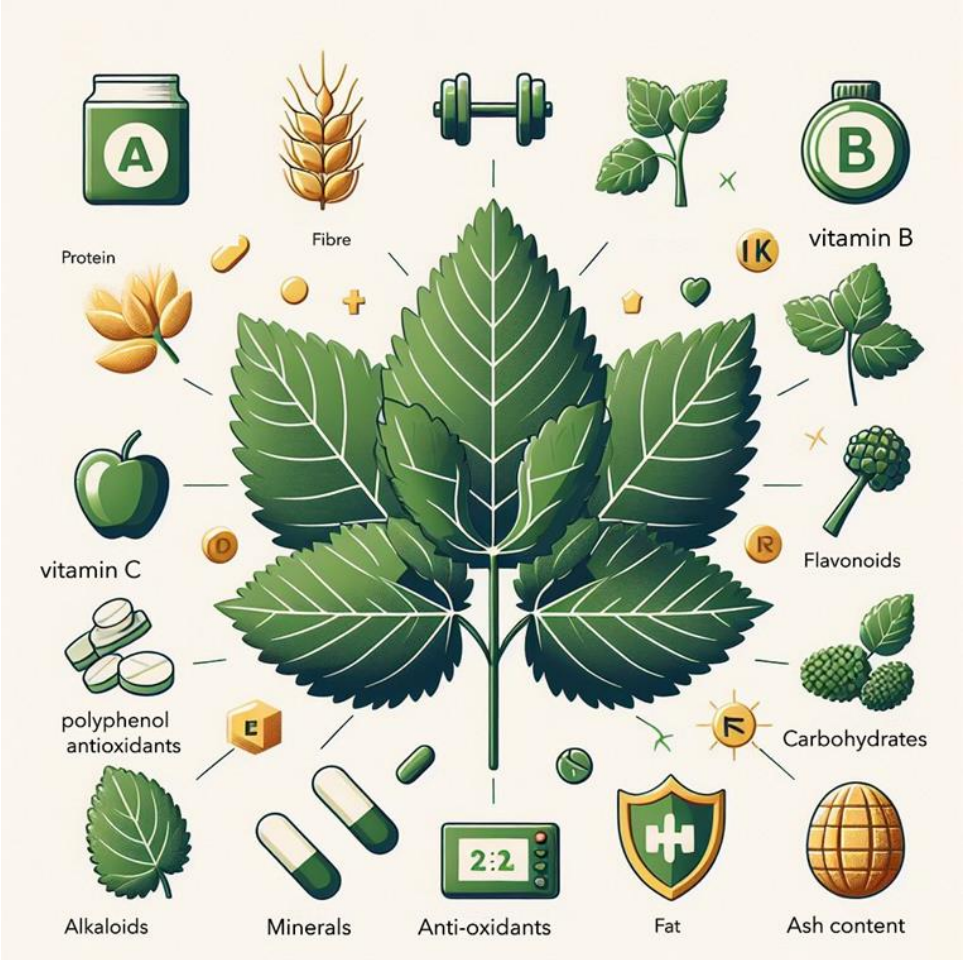


Figure 1: Nutritional Content of Mulberry Leaves

1.6 IMPORTANCE OF *MORUS INDICA*

1.6.1 MEDICINAL IMPORTANCE

The fast-growing perennial plant *Morus indica* (L.), commonly known as Indian mulberry, belongs to the *Moraceae* family and serves as the primary food source for the monophagous insect *Bombyx mori*, making it highly relevant in sericulture [41,42]. Ancient Ayurvedic texts document its therapeutic attributes, utilizing its fruits, leaves, roots, bark, and latex in diverse formulations aimed at addressing various ailments. The presence of a plethora of bioactive compounds, including steroids, phenolics, alkaloids, reducing sugars, tannins, flavonoids, anthocyanins, terpenoids, cardiac glycosides, saponins, proteins, amino acids, and carbohydrates, underscores its traditional use as a natural remedy for antibacterial, antifungal, anti-hyperglycemic, and anti-hypolipidemic conditions.

Recent studies, supported by citations [43] through [44], have unveiled the antibacterial prowess of

Morus indica. Compounds like Kuwanon C and G, identified in the root bark extract, exhibit notable antibacterial activity. Particularly strong antibacterial effects are observed in hydro methanolic stem bark extracts, suggesting their potential as effective antimicrobial agents against gastrointestinal illnesses [45].

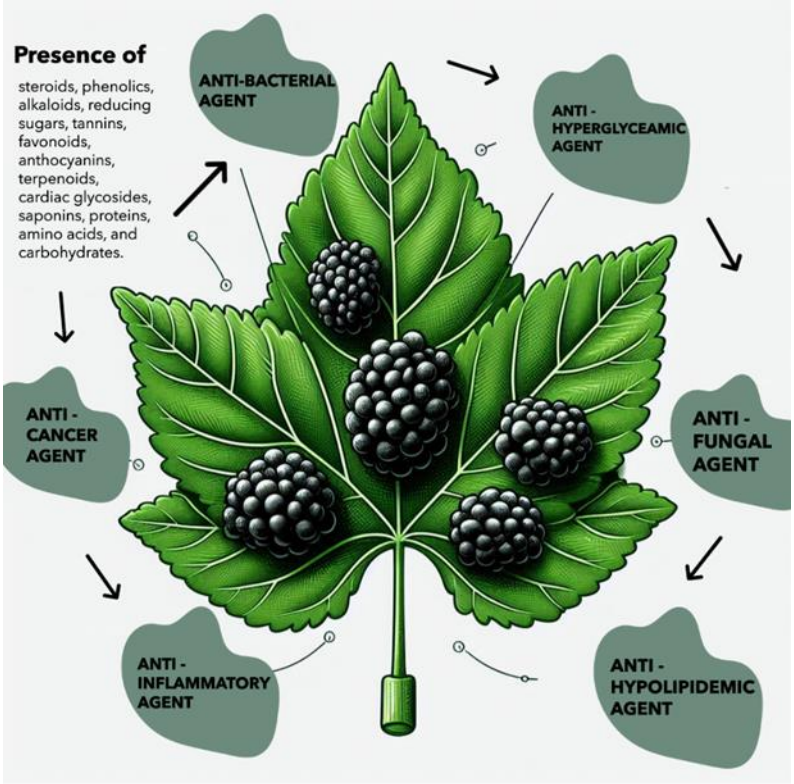


Figure 2: Medicinal Importance of Mulberry

Moreover, *Morus indica* leaves have been found to possess antibacterial properties effective against oral microbes [46]. Phytochemicals within *Morus indica*, such as phenolic compounds, flavonoids, and anthocyanins, contribute to its antimicrobial capabilities. Further research, as detailed in citations [47] through [48], has demonstrated *Morus indica's* effectiveness against various bacterial species, including oral pathogens. Its ability to impede DNA replication, metabolic processes, and disrupt cell membranes contributes to this efficacy. Additionally, its antibacterial properties are augmented by its capacity to neutralize free radicals and alleviate oxidative stress, thereby mitigating damage to bacterial cells. Moreover, *Morus indica* exhibits anti-inflammatory properties due to the presence of active compounds such as phenolic compounds, flavonoids, and anthocyanins. These compounds contribute to its anti-inflammatory effects, offering potential applications in combating various inflammatory conditions. Inhibition of inflammatory mediators: The active ingredients in *Morus indica* can decrease inflammation by preventing the generation of inflammatory mediators including prostaglandins and cytokines [49]. Free radicals can produce oxidative stress, which can lead to inflammation. *Morus indica* possesses antioxidant activity that can help avoid this.

1.6.2 COMMERCIAL IMPORTANCE OF *MORUS INDICA*

The commercial value of Indian mulberry, also known as *Morus indica*, is significant, particularly within the sericulture sector where it serves as the primary food source for the monophagous insect *Bombyx mori* [50]. Mulberry's impact extends to the rural economy [51]. *Morus indica* fruit shows promise for inclusion in food compositions due to its abundance in bioactive compounds and its anti-quorum sensing properties [52]. Additionally, advancements such as an effective in vitro regeneration process enable its potential for genome editing and genetic modification [53]. The acquisition of a high-quality genome sequence for *Morus indica* provides a comprehensive resource for genomic research and agricultural advancement. The bioactive components of Indian mulberry have been harnessed for their medicinal properties, suggesting potential applications as nutraceuticals for various health concerns. Moreover, mulberry finds its place in numerous cosmetic products and functional foods [54-56]. The plant's significance extends to the food and pharmaceutical industries [57]. The potent antioxidant activity observed in mulberries hints at their potential therapeutic applications, including in conditions like cancer and cardiovascular disease. They also show promise in preventing fatty liver disease and managing disorders linked to hyperlipidemia by aiding in lipid reduction [58]. *Morus indica's* commercial importance within the sericulture sector remains paramount, given *Bombyx mori's* heavy reliance on it. The demand for products derived from mulberries, boasting beneficial biological effects, underscores its significance in various industries. Mulberry plays a crucial role in the global economy and holds multifaceted importance across diverse sectors [53-57].

1.6.3 ECONOMIC IMPORTANCE OF MULBERRY (*MORUS INDICA*) IN INDIA

Mulberry (*Morus indica*) is a cornerstone of India's economic landscape, playing a crucial role across sectors such as sericulture, agriculture, horticulture, and traditional medicine. Its multifaceted contributions sustain livelihoods, foster income generation, drive exports, and promote sustainable agricultural practices, establishing it as an indispensable asset for economic growth and rural development in the country.

The sericulture industry, deeply entrenched in Indian tradition, relies heavily on mulberry leaves as the primary source of nutrition for silkworms. This sector not only offers employment opportunities to rural communities but also holds significant sway over the nation's economy. India ranks among the largest silk producers globally, with states like Karnataka, Andhra Pradesh, and Tamil Nadu emerging as major silk-producing regions. The sericulture industry, encompassing silk farming, cocoon production, and silk processing units, generates substantial income for a significant portion of the population [58].

Beyond silk production, mulberry trees play a vital role in Indian agriculture, benefiting sectors like traditional medicine, horticulture, and livestock farming. With its diverse contributions supporting

sustainable agricultural practices, export growth, income generation, and livelihoods, *Morus indica* holds a pivotal and multifaceted economic function in India. The nation's economy heavily relies on its contributions across various industries, enhancing the quality of life for millions of people. The sericulture sector, a longstanding and successful industry in India, predominantly feeds silkworms on mulberry leaves, contributing significantly to the nation's economy while offering employment prospects in rural areas. Mulberry leaves, nutrient-rich fodder, also benefit the dairy and animal husbandry industries. Moreover, mulberry cultivation aids in sustainable farming methods by enhancing soil fertility and reducing erosion. Practices like mulberry crop rotation further bolster overall agricultural production, further solidifying its importance in India's economic landscape.

Agricultural Significance: Some mulberry species are farmed for their juicy, flavorful fruits in India. In addition to being eaten raw, these fruits are often used to produce jams, jellies, and confections. Selling mulberry fruits is one way for farmers in mulberry-growing regions like Punjab and Haryana to augment their revenue. The nursery and landscaping industries gain from the ornamental attractiveness of mulberry trees [59].

Traditional medicine and herbal remedies: The leaves, bark, and fruits of the mulberry tree are employed for their curative characteristics in Indian traditional medicine systems like Ayurveda. Due to the expanding demand for natural and conventional treatments, the manufacturing and marketing of herbal medications and nutritional supplements containing components obtained from mulberries has grown significantly in economic importance. The mulberry tree has a significant economic impact on India. Employment possibilities are created as a result, particularly in rural regions where sericulture is practiced extensively. Along with silk production, other income sources include selling mulberry fruits, raising cattle, and expanding herbal medicine market. India's involvement in the international silk market helps the country earn foreign currency and expand its economy [60].

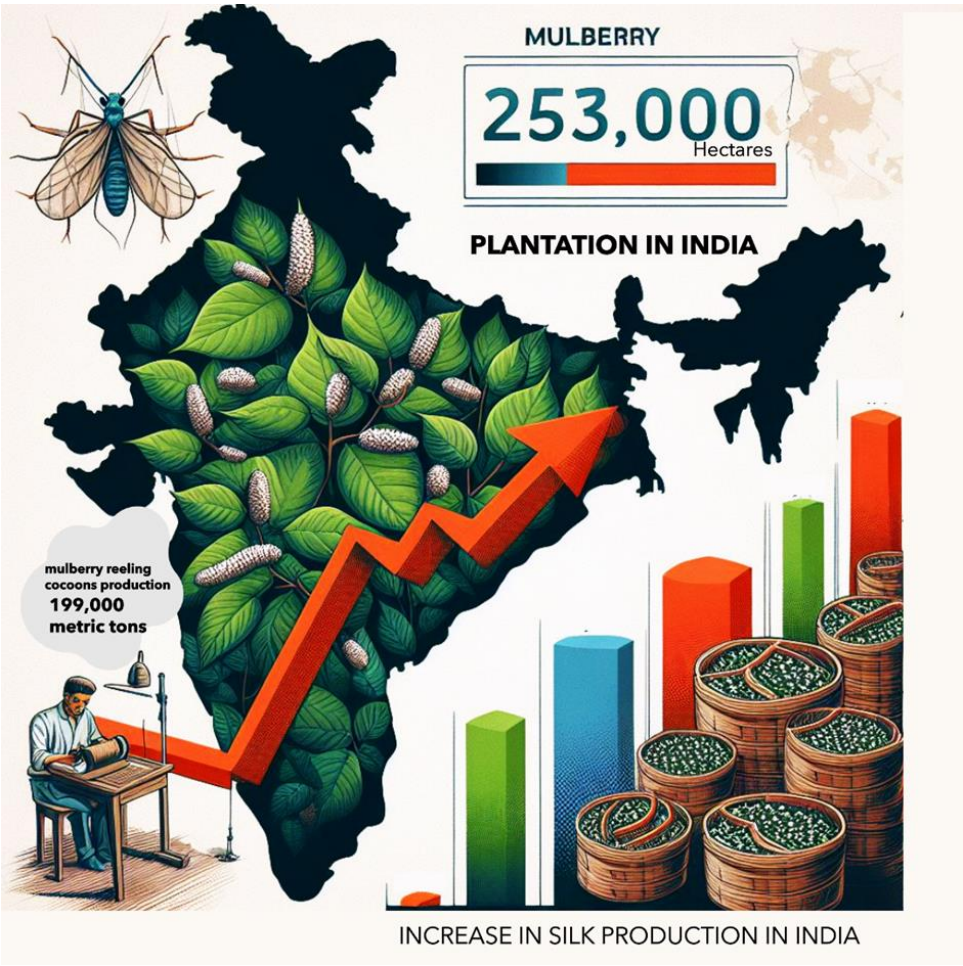


Figure 3: Mulberry Production and Silk Industry in India

1.7 BREEDING TECHNOLOGIES OF MULBERRY

Mulberry breeding methods have evolved significantly, encompassing various techniques to enhance the yield, quality, and resilience of this versatile plant. It takes many years (approximately 15-20) to develop a new variety of mulberry because it is a perennial woody plant. Breeding targets should be set with a long-term view. To date, breeding targets have been, for example, high yield, high nutritional value and resistance against diseases and pests. But today, new targets have been added to cope with changes in the sericultural system, such as large numbers of silkworm reared and adaptability to densely planted fields suitable for mechanical harvesting. Crossing is the major breeding method adopted for the development of new mulberry varieties. The aim was to develop a variety with high quality, high yield and resistance Conventional breeding involves hybridization and selection, where different mulberry varieties are crossed to create offspring with desired traits. These offspring are rigorously screened and bred further to establish new varieties with improved characteristics [61]. Mutagenesis exposes mulberry plants to mutagenic agents, like radiation or chemicals, to induce DNA mutations. Subsequent screening identifies plants with favorable traits for further breeding. Molecular breeding

relies on molecular markers to select plants based on their DNA, rather than physical traits, aiding in optimizing mulberry leaf yield and quality.[59] Plant tissue culture fosters disease-free and trait-specific plant propagation in a sterile environment, facilitating mass production of desired mulberry traits. Genetic engineering manipulates mulberry DNA to enhance leaf yield and quality, as well as abiotic stress tolerance [60]. Collectively, these breeding technologies illustrate the promising future of mulberry farming, driven by continuous advancements in breeding techniques [62]. Mulberry propagation is generally carried out by grafting and by cutting methods. Root grafting prevails because it is easy to handle and the grafted saplings have a high survivability. The cutting method can be with hard wood (using the branches grown in the previous year) and soft wood (using the spring sprouted shoots). With mulberry varieties of poor rooting ability, treatment with plant hormones is advised to stimulate rooting. Recently, tissue culture derived saplings have also been produced [63].

1.8 GENOMIC RESOURCES OF MULBERRY

The need for mulberry genetic resources has been acknowledged in the expansion of the mulberry industry and sustainable silk production [64,65]. Research endeavors have focused on assessing genetic diversity and variation among mulberry genotypes to enhance leaf productivity and other desirable traits. Modern methodologies like inter simple sequence repeat (ISSR) analysis have been employed to identify genetic variations among mulberry genotypes from diverse regions [60]. Preservation and utilization of wild mulberry genetic resources, which harbor genes for crucial traits like drought and salt resistance, have been emphasized [66].

Utilizing biotechnological approaches such as molecular markers, cloning of essential functional genes, and transgenic experiments, the genetic resources of mulberry have been explored, characterized, conserved, and utilized. These efforts significantly contribute to advancing knowledge and leveraging mulberry genetic resources for sustainable sericulture and crop improvement. Given its significance as a major food source for silkworms, mulberry (Morus spp.) holds considerable importance in India [67]. Recent interest in mulberry genomics stems from its potential to offer valuable resources for functional and translational genomics. The high-quality draft genome sequence of the Indian mulberry cultivar K2 (*Morus indica*), with significant repetitive content, stands out as a notable genetic resource for mulberries in India [68]. This genome serves as a crucial asset for research in functional and translational genomics. Advances in mulberry genetics and genomics have been facilitated by sequencing the draft genome of *Morus notabilis* and various transcriptome and genomic resources. Studies on mulberry genome size, genetic diversity, and phenotypic variation have provided insights into genetic differences and their impact on trait plasticity [69].

The synthesis of the first T2T gap-free reference genome of mulberry, enabled by integrating high-coverage, precise long-read sequences, offers a vast resource for exploring the structure and

development of polycentric chromosomes. Furthermore, ISSR research has contributed to assessing genetic divergence across mulberry cultivars and variations in India, supporting effective conservation and utilization of mulberry genetic resources [70]. Research on Arabidopsis SHN1 expression in Indian Mulberry (*Morus indica* L.) has furnished insights into genetic engineering techniques aimed at enhancing leaf surface wax content and reducing post-harvest water loss.

In summary, India's rich and invaluable mulberry genomic resources present opportunities for in-depth exploration of mulberry genetics and genomes, promising significant contributions to functional and translational genomics efforts [71]. Draft genomic sequence of *Morus indica*

The Indian mulberry's (*Morus indica*) draught genome sequence is a complete tool for research in functional and translational genomics. Multiple methods were used to generate the genome sequence, which produced a high-quality sequence with a gene completeness of 96.5 percent. [72] The genome study of M. indica revealed many protein-coding genes, repetitive DNA sequences, and species-specific genes [73]. Understanding the biological processes in mulberry was made possible by transcriptome profiling, which identified tissue-specific and differential gene expression [74]. Morus notabilis, a species of wild mulberry, had its genome sequence examined as well, and a web-based database called MorusDB was created to make it simple to access mulberry genetic data [75]. The discovery of growth-regulating factors (GRFs) in mulberries advances our knowledge of how this significant crop grows and develops [76]. Earlier work on the whole-genome sequencing of *Morus indica* was completed by the School of Computational & Integrative Sciences at Jawaharlal Nehru University in New Delhi. This served as the foundation for our efforts, with over 90% of them being confirmed by transcript data, the genome sequence analysis found 27,435 high-confidence protein-coding genes and 49.2% of repetitive DNA. 4.8% of the genes in the M. indica genome were found to be species-specific when compared to other plant genomes. Across numerous accessions, transcriptome profiling revealed tissue-specific and variable expression of protein-coding genes, with around 4.7 percent and 2–5 percent of genes implicated in various biological activities. Following the whole genome resequencing of 21 accessions/species, 2.5 million single nucleotide polymorphisms and 0.2 million insertions/deletions were found. The information and findings from this study served as a thorough resource for our mulberry genomes research and its advancement [77].

1.9 EXPLORING THE GENOME OF *MORUS INDICA*

1.9.1 WHOLE GENOME SEQUENCE OF MULBERRY *MORUS INDICA*

27,435 high-confidence protein-coding genes were found in the genome sequencing study, with 49.2% of them having repetitive DNA, and more than 90% of them having transcript evidence. The M. indica genome contains 4.8 percent of species-specific genes, according to comparison with other plant genomes. Across numerous accessions, transcriptome profiling revealed tissue-specific and variable expression of protein-coding genes, with around 4.7 percent and 2–5 percent of genes implicated in various biological activities. A total of 2.5 million single nucleotide polymorphisms and 0.2 million insertions/deletions were found during whole genome resequencing of 21 accessions/species. Our genomics research on mulberry and its advancement benefited greatly from the information and findings from this work. [77]

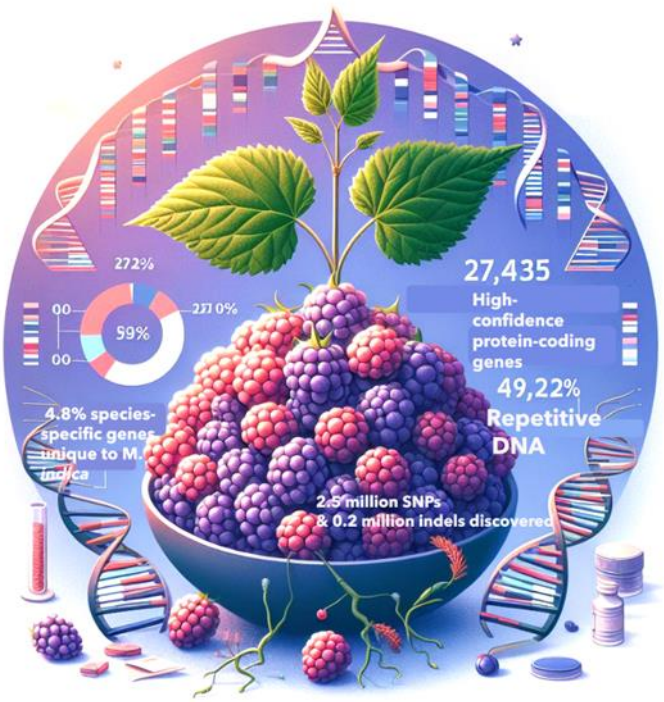


Figure 4: Genome information for *Morus indica*

1.9.2 THE REPETITIVE CONTENT IN A DRAFT GENOME

In many biological processes and genome evolution, repetitive sections of the genome are significant. They can influence structural variation, genomic stability, and gene regulation. The repeated sections in the Indian mulberry genome study were found using a mix of data from many distinct methods, including Illumina sequencing, single-molecule real-time sequencing, chromosomal conformation capture, and optical mapping. Repetitive DNA makes about 49.2% of the Indian mulberry genome, according to the research. It is crucial to identify repeating sections because they can provide light on

how the genome is organized and structured and may also include functional components. The annotation and characterization of genes, as well as the investigation of genome evolution and genetic diversity, can all benefit from an understanding of the repetitive areas. The Indian mulberry genome has repeated sections, which is relevant information for future functional and translational genomics studies of this significant agricultural plant [78].

1.9.3 GENES THAT CODE FOR PROTEINS

In a number of investigations, the farmed Indian mulberry *Morus indica's* protein-coding areas have been found and described. 27,435 high-confidence protein-coding genes were found in the M. indica genome, and transcript evidence backed up more than 90% of them [79]. These genes are involved in tissue-specific and varied expression across several accessions, and they are essential for a variety of biological functions. Furthermore, a cDNA library made from M. indica revealed a significant number of genes encoding enzymes involved in diverse secondary metabolites, opening up new study directions for mulberry natural product chemistry. The role that protein-coding regions play in plant stress responses and cellular metabolism makes them significant. The research Centre of Jawaharlal Nehru University carried out [80]. Using the repeat-masked genome sequence, a mix of de novo prediction, similarity-based searches, and evidence-based searches were used to identify the protein coding regions in the Indian mulberry genome. 39 Mb (7.7%) of the genome's sequence included 27,435 unique genes that the investigation anticipated. Multiple transcript isoforms were predicted for at least 1546 genes [81]. The predicted genes' average mRNA length was 1430 bp. With an average exon count of 6, almost 87 percent of the genes were expected to be multi-exonic. Based on similarity searches in several databases, functional annotation of predicted genes was carried out. In all, 27,138 genes (98.9%) showed notable sequence similarity with proteins included in the UniProt Viridiplantae database. Additionally, it was discovered that 24,939 (90.9%) of the genes included at least one conserved Pfam domain. Through whole-genome resequencing of 21 mulberry accessions/species, it was possible to identify DNA polymorphisms and create a high-density SNP map. This found 2.5 million single nucleotide polymorphisms and 0.2 million insertions/deletions [82].

1.9.4 IDENTIFICATION OF SINGLE NUCLEOTIDE POLYMORPHISM IN *MORUS INDICA*

The draught genome contains comprehensive information on the SNPs found in *Morus indica*. To find SNPs in *Morus indica*, the whole genomes of 21 mulberry accessions/species were sequenced. High-quality reads were filtered out of the raw data, representing extremely high coverage of the mulberry genome for each genotype. [83] A total of 95.3-99.7% of high-quality reads were successfully mapped to the M. indica draught genome assembly. 2,480,073 high-quality SNPs and 2,14,704 insertions and deletions (InDels) were found in total. SnpEff was used to annotate DNA polymorphisms in coding

areas as synonymous, non-synonymous, and large-effect (v4.3t). Boxplot analysis was used to estimate the frequency (per kb) distribution of SNPs and InDels in the coding and 2 kb upstream areas, and the genes with substantially higher frequency of SNPs/InDels above the upper fence in the boxplot were selected. Utilizing Cytoscape, comparative GO enrichment maps for various gene sets were also created. [84] Numerous applications, including genetic mapping, association research, and population genetics, depend on the detection of SNPs in the *Morus indica* plant. [85] SNPs can be used as markers for genetic mapping to locate the genes responsible for phenotypes. They can also be used to identify genetic variants associated with complex traits or diseases in association studies. In population genetics, Studying the genetic diversity and population organization using SNPs can reveal information about the evolutionary background and adaption of a species [86] Whole-genome resequencing of 21 mulberry accessions/species was used to discover SNPs before high-quality reads were filtered and mapped to the draught M. indica genome assembly. Utilizing tools like SnpEff and Cytoscape, the SNPs were annotated and examined for frequency distribution and gene enrichment.

An important resource for functional and translational genomics studies aimed for mulberry improvement is the discovery of SNPs in *Morus indica*. SNPs may be employed to examine the genetic diversity and structure of mulberry populations, generate molecular markers for breeding plans, and discover genes linked to desired features [87].

1.9.5 SIMPLE SEQUENCE REPEATS

Microsatellites, also known as simple sequence repeats (SSRs), are short repetitive DNA sequences scattered throughout the genome. These SSRs serve as important genetic markers with diverse applications such as marker-assisted selection, population genetics, and genomic mapping. Utilizing a combination of technologies including Illumina, PacBio single-molecule real-time sequencing, and Bionano optical mapping, the draft genome sequence of *Morus indica* was analyzed to identify SSRs. Within the *M. indica* genome, 121,212 SSRs were identified, occurring at an average frequency of one SSR per 3.2 kb. The most common SSR motifs included mononucleotide (57.5%), dinucleotide (28.7%), trinucleotide (11.5%), tetranucleotide (1.7%), pentanucleotide (0.4%), and hexanucleotide (0.4%) repeats. These SSRs serve as molecular markers for genetic mapping and population genetics research.

Repetitive regions within the *Morus indica* genome were identified using various tools such as RepeatMasker and RepeatModeler to locate SSRs. The distribution frequency and different motif types of the identified SSRs were annotated and analyzed. These SSRs offer diverse utilities including marker-assisted selection, population genetics, and genomic mapping. They enable the examination of genetic diversity and population structure in mulberry populations, aid in the discovery of genes associated with desirable traits and facilitate the development of molecular markers for breeding

programs. The exploration of SSRs in *Morus indica* holds significant promise for advancing functional and translational genomics in mulberry improvement. These SSRs offer avenues to investigate the genetic diversity and structure of mulberry populations, identify genes associated with desired traits, and develop molecular markers for breeding strategies. The draft genome sequence of *Morus indica*, generated through a fusion of four distinct technologies—namely, Illumina, PacBio single-molecule real-time sequencing, chromosomal conformation capture (Hi-C), and Bionano optical mapping—served as the foundation for SSR discovery [88].

1.10 PHENOTYPES OF MULBERRY STUDIES IN THIS RESEARCH

The production of mulberries (Morus spp.) is crucial to the sericulture sector since they serve as the main food supply for silkworms (Bombyx mori), which in turn helps the silk manufacturing sector. This study begins a thorough investigation of mulberry phenotypes, including drought stress tolerance, leaf yield, root rot resistance, and nitrogen consumption efficiency. The goal is to further knowledge of mulberry genetics and aid in the creation of improved mulberry cultivars. Due to their crucial roles in mulberry farming and their capacity to completely alter sericulture techniques, t phenotypic features have been chosen with consideration.

1.10.1 PROBLEMS ENCOUNTERED BY MULBERRY CULTIVATORS

The above-mentioned conditions frequently present mulberry farmers with a number of difficult problems, such as drought stress, root rot infections, and insufficient nitrogen usage efficiency. Reduced water availability brought on by drought stress can result in lower mulberry leaf production and, as a result, a reduction in the amount of food available to silkworms. In especially in areas susceptible to water constraint, this has a direct influence on the silk production cycle and, eventually, the revenue of sericulture producers [89]. Diseases that cause root rot are still another serious issue. These diseases, which cause wilting, leaf discoloration, and decreased plant vigor, can completely destroy mulberry trees. The ensuing fall in leaf yield puts the nutritional requirements of silkworms in danger, perhaps resulting in less silk being produced. Additionally, the need for chemical treatments to battle root rot infections adds to farmers' expenses and might have a negative impact on the environment. The difficulties experienced by mulberry producers are further exacerbated by suboptimal nitrogen usage efficiency [90]. Ineffective nitrogen absorption and use can result in higher fertilizer usage, rising production costs, and nitrogen runoff that contributes to environmental damage [91]. Diseases that cause root rot are still another serious issue. These diseases, which cause wilting, leaf discoloration, and decreased plant vigor, can completely destroy mulberry trees. The ensuing fall in leaf yield puts the nutritional requirements of silkworms in danger, perhaps resulting in less silk being produced. Additionally, the need for chemical treatments to battle root rot infections adds to farmers' expenses and might have a negative impact on the environment. The difficulties experienced by mulberry

producers are further exacerbated by suboptimal nitrogen usage efficiency. [80] Ineffective nitrogen absorption and use can result in higher fertilizer usage, rising production costs, and nitrogen runoff that contributes to environmental damage [92]. Farmers face a complicated conundrum that highlights the need of tackling nitrogen management in mulberry farming as they struggle with the combined weight of economic inefficiency and environmental responsibility. Together, these problems highlight the complex interactions between environmental stresses and agricultural production in the mulberry industry. In order to successfully navigate these difficulties, farmers must strike a balance between the need for increased yield, cost-effective methods, and sustainable resource management, all the while protecting the silk production process, which is closely related to the availability and quality of mulberry leaves. For the sericulture industry's long-term viability and the industry's overall environmental effect, as well as the financial well-being of farmers, it is imperative to address these issues via research and innovation.

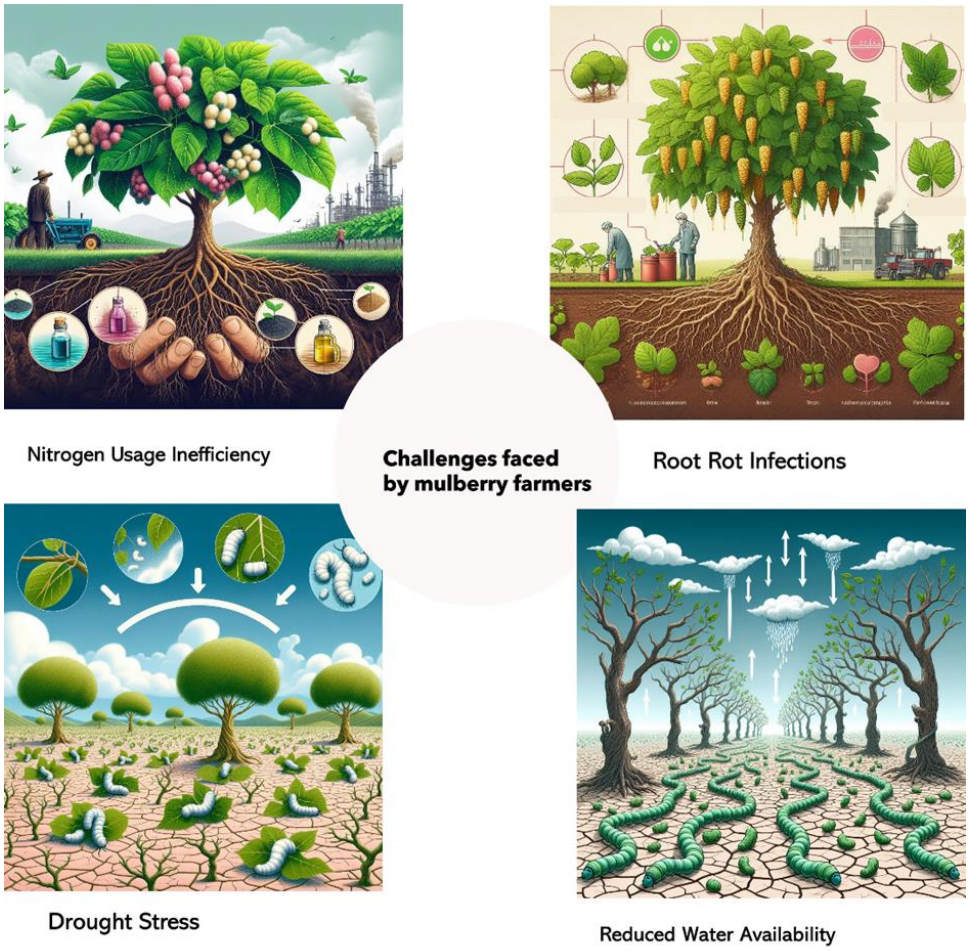


Figure 5: Illustration of the various challenges faced by mulberry farmers

1.10.2 DROUGHT STRESS

The primary focus of this research revolves around the issue of drought stress, given its significant implications for mulberry production. The escalation in the frequency and severity of drought events, attributed to climate change and erratic weather patterns, underscores the urgency of understanding mulberry's genetic resilience to drought stress. This comprehension holds immense importance as mulberry's ability to endure drought conditions is pivotal for sustaining sericulture production, especially in regions with limited water resources [93]. The exacerbation of water scarcity due to climate change could disrupt traditional sericulture practices, highlighting the necessity for drought-tolerant mulberry cultivars. Thus, this study aims to identify the genes and mechanisms underlying drought tolerance in various mulberry varieties to develop cultivars capable of thriving in arid environments. Beyond its implications for sericulture, mulberry farming plays a crucial role in agroforestry, providing livelihoods for local communities in numerous regions. By cultivating drought-resistant mulberry cultivars, farmers can mitigate the impact of environmental uncertainties on their income, reducing their vulnerability to crop failures [94]. Rigorous phenotyping methods will be employed to evaluate mulberry plants' response to water scarcity, examining variables such as stomatal conductance, relative water content, and leaf water potential.

1.10.3 LEAF YIELD

Leaf yield stands as a pivotal trait in this study, bearing significant importance not only for sericulture but also for agriculture. It directly correlates with the quantity of mulberry leaves a plant produces, playing a crucial role in determining the output of silkworms (Bombyx mori) [95]. The essence of leaf yield lies in its direct impact on silkworms' development, silk production, and overall well-being, given that mulberry leaves serve as their primary food source [96]. Thus, it becomes imperative for the sericulture sector to unravel the genetic components responsible for high leaf production. High leaf yield offers a multitude of advantages. Firstly, it ensures a consistent and ample supply of food for silkworms, thereby accelerating their growth rates and silk output. Ultimately, the aim of sericulture is to enhance silk cocoon production, a goal attainable through increased leaf yield. Moreover, a reliable and high-quality leaf supply reduces stress among silkworms, bolstering their resilience and overall health. Financially speaking, high leaf production benefits sericulture producers by optimizing silk production capacity and increasing revenue. Furthermore, it may extend the silk-producing season, enabling growers to capitalize on market demands [97].

Technologies employed for evaluating leaf yield typically involve non-destructive techniques, given the necessity to retain leaves for ongoing silkworm feeding. Methods such as leaf area and dry weight measurements are commonplace, with image analysis software often utilized to estimate leaf area by determining the total leaf surface area. Harvested leaves are dried to remove moisture before being

weighed to estimate their dry weight, offering quantitative insights into leaf production without causing harm to the plant. The study of leaf yield holds significance beyond sericulture, as mulberry cultivars with high yields can enhance overall agricultural production. Increased leaf output can serve as additional green fodder for animals or organic mulch in agroforestry systems, enriching soil fertility and crop health. Mulberry leaves, being a rich source of nutrients, can also be utilized in various leaf-based products such as tea or herbal treatments. Ultimately, leaf yield emerges as a crucial trait with far-reaching implications for agriculture and sericulture alike [88]. Understanding the genetic factors contributing to high leaf yield facilitates the development of mulberry cultivars that not only boost silk production and enhance financial returns for sericulture farmers but also potentially contribute to broader agricultural sustainability through increased fodder and nutrient-rich leaf production [98].

1.10.4 NITROGEN USE EFFICIENCY

Exploring Nitrogen Use Efficiency (NUE) within the context of mulberry cultivation holds paramount importance due to its profound implications for agriculture and sustainability. NUE plays a pivotal role in maximizing agricultural productivity, efficient resource utilization, and environmental preservation. Nitrogen stands as a vital nutrient essential for plant growth and development, making the analysis of NUE a crucial and intricate aspect of mulberry farming [99]. Given its far-reaching consequences, NUE assumes critical significance in the realm of mulberry cultivation. NUE is relevant to mulberry cultivation as it directly impacts crop yield, influencing silkworm productivity, a key factor in sericulture. Nitrogen, being a crucial macronutrient, plays a pivotal role in promoting plant growth and leaf formation, rendering mulberry leaves an ideal food source for silkworms. Higher NUE signifies the ability of mulberry plants to enhance leaf production while efficiently utilizing available nitrogen resources, ensuring a consistent and abundant supply of premium mulberry leaves. This, in turn, enhances silkworm health and production, thereby contributing significantly to the sustainability of the sericulture sector [100].

Furthermore, investigating NUE in mulberry cultivation facilitates cost-effective resource management and environmental sustainability. Inefficient nitrogen utilization can lead to financial losses, given that nitrogen fertilizers constitute a significant input cost in agriculture. Precision nitrogen management, facilitated by NUE research, reduces fertilizer usage, conserves resources, and mitigates the adverse environmental impacts of nitrogen runoff, soil acidification, and groundwater contamination. This fosters a more responsible and environmentally friendly approach to mulberry farming, aligning with principles of sustainable agriculture and environmental stewardship. Moreover, NUE research addresses broader concerns related to agricultural sustainability and climate resilience. Nitrogen pollution, largely attributable to agricultural activities, exacerbates numerous environmental issues. By improving NUE, mulberry cultivation can mitigate nitrogen losses into the environment, thereby

mitigating problems like eutrophication and air pollution. Additionally, effective nitrogen utilization becomes crucial for crop resilience in the face of climate change-induced uncertainties in temperature and precipitation patterns. High NUE mulberry cultivars enable sericulture to adapt to climatic challenges by thriving in changing environmental conditions [101].

1.10.5 ROOT ROT CONDITIONS

This study centers on an in-depth examination of the root rot fungal condition affecting mulberry (Morus spp.), addressing a significant challenge in mulberry cultivation. Root rot infections, caused by various pathogenic fungi, pose a serious threat to the health and productivity of mulberry plants. A detailed exploration of this issue underscores the urgent need to expand our understanding of root rot fungal infections, their implications, appropriate assessment methodologies, and the broader context of prior research in this domain. Effective management of these diseases necessitates a comprehensive understanding of the root rot fungal condition. These infections manifest through symptoms such as wilting, leaf yellowing, root discoloration, and overall weakened plants. By identifying the specific fungal pathogens responsible for root rot conditions through visual inspections and laboratory analyses, targeted treatments can be devised to mitigate their impact and prevent crop losses. Given that mulberry leaves serve as the primary food source for silkworms (Bombyx mori) in sericulture, addressing root rot infections is crucial for ensuring a consistent supply of high-quality mulberry leaves, thereby promoting silkworm health and production, essential for the sustainability of the sericulture sector. Furthermore, addressing the root rot fungal issue in mulberry plants aligns with ecologically friendly agricultural practices. Developing mulberry cultivars that naturally resist these infections obviates the need for synthetic fungicides, contributing to environmentally responsible farming practices. By investigating the genetic basis of resistance, this study holds the potential to develop robust and durable mulberry cultivars, particularly relevant amid the imperative for responsible resource management in agriculture and changing environmental conditions. The evaluation of the root rot fungal condition employs a multimodal approach, encompassing visual observation of disease symptoms, isolation and identification of fungal pathogens from infected plant samples using microscopy and molecular techniques, and assessment of disease severity. Advanced molecular techniques such as DNA sequencing and PCR enable precise identification and classification of root rot pathogens, shedding light on the diversity of fungi affecting mulberries. Drawing on prior research in the field of root rot diseases, including studies on pathogenicity mechanisms, host specificity, and management strategies across various crops, this research expands our understanding of the unique challenges faced by mulberry growers and silk producers. It builds upon this foundation to foster the resilience of the sericulture industry over the long term by developing disease-resistant mulberry varieties.

In conclusion, investigating the root rot fungal condition in mulberry cultivation is a critical endeavor with implications that extend beyond crop health and productivity, encompassing the sustainability of agriculture and the prosperity of the sericulture sector.

1.11 IDENTIFICATION OF BIOMARKERS FOR THE FOUR PHENOTYPES OF MULBERRY

The principal discovery of this study, with significant implications for sericulture and sustainable agriculture, involves the identification of biomarkers associated with leaf yield, root rot resistance, and nitrogen utilization efficiency in mulberry farming. Biomarkers serve as physiological or molecular indicators linked to the presence or severity of a particular trait or condition. It is imperative to identify these biomarkers due to factors such as precision breeding, crop management, and the development of robust, disease-resistant mulberry cultivars. They can function as early indicators of stress and plant health, particularly in plants capable of withstanding drought stress, facilitating proactive management. These biomarkers may include physiological indicators such as leaf water potential or relative water content, or they may be specific gene expressions associated with stress response pathways. Technologies like RNA-Seq can be employed to identify candidate genes whose expression patterns correlate with drought tolerance, shedding light on the molecular mechanisms underlying this trait. Previous research in crops like maize and rice has unveiled drought-related biomarkers that can inform mulberry studies. Biomarkers offer a means to anticipate and select high-yielding mulberry cultivars during breeding programs focusing on leaf production. They could be based on genes that enhance leaf yield or linked to critical physiological processes influencing leaf development. Molecular approaches such as transcriptomics and marker-assisted selection can be utilized to discover and validate these biomarkers. Although research on mulberry-specific leaf yield biomarkers is limited, insights from studies on other crops provide valuable information on the genetic basis of high-yielding traits. Identifying mulberry plants with innate resistance to root rot pathogens necessitates the use of biomarkers for root rot resistance. These biomarkers may encompass biochemical indicators associated with pathogen interactions, disease resistance mechanisms, or genes involved in defense responses. Genomic methods like genome-wide association studies (GWAS) and molecular techniques for identifying disease resistance genes can aid in identifying these biomarkers. Previous studies on other crops have identified biomarkers linked to resistance against various diseases, offering guidance for identifying similar biomarkers in mulberry. Efficient nitrogen management in mulberry agriculture relies on indicators for nitrogen utilization efficiency. These biomarkers may include physiological signals indicating nitrogen status in the plant and genes involved in nitrogen absorption, assimilation, or transport. Molecular methods such as gene expression analysis and metabolomics can help discover

nitrogen-related biomarkers. While research on mulberry-specific biomarkers for nitrogen utilization efficiency is nascent, lessons from analogous studies in crop species like rice and wheat can inform the search for relevant biomarkers in mulberries. The discovery of biomarkers for these traits in mulberry represents a cutting-edge endeavor with the potential to enhance crop management, breeding programs, and sericulture practices. These biomarkers serve as valuable tools for meticulous selection, improved crop management, and the development of resilient and high-yielding mulberry cultivars. Despite potential limitations in prior research on biomarkers for these specific traits in mulberry, the wealth of information from analogous studies in other crops provides a solid foundation for this crucial research endeavor [102].

1.12 GENETIC LINKAGE & MAPPING STUDIES

Genetic linkage mapping is an indispensable method in genetic studies that delineates the locations of genes and genetic markers on chromosomes, as well as their proximity. This technique holds immense importance for unraveling the genetic underpinnings of various traits, such as yield, disease resistance, and stress tolerance, within the context of mulberry (*Morus indica*), a tree species of significant economic importance for the sericulture industry. Introducing linkage mapping in mulberry is imperative to advance breeding programs, enhance crop management practices, and augment overall mulberry productivity [103].

The significance of mulberry linkage mapping lies in its provision of a blueprint of the mulberry genome and identification of genes and markers associated with desirable traits. This genetic roadmap facilitates the improvement of mulberry cultivars through selective breeding, yielding varieties that are high-yielding and resistant to diseases. Moreover, linkage mapping aids in the development of sustainable sericulture methods by elucidating the genetic pathways underlying complex traits like drought tolerance, leaf yield, and nitrogen use efficiency. Understanding the genetic structure of mulberry enables researchers and breeders to expedite the creation of robust and productive cultivars, essential for the expansion of the silk industry.

Mulberry linkage mapping can be conducted using various methods, with molecular markers being the primary tools. Commonly employed markers for linkage mapping encompass simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs), and amplified fragment length polymorphisms (AFLPs) [104]. These markers are selected based on their genome-wide distribution and polymorphic characteristics. Typically, genetically diverse mulberry individuals are crossed, and the resulting progeny are genotyped using molecular markers. Subsequently, data from this genotyping is processed using specialized software and statistical techniques to unveil marker-trait associations and construct genetic maps.

In summary, linkage mapping in the mulberry species *Morus indica* serves as a crucial genetic research

tool with significant implications for sustainable agriculture and the sericulture sector. By elucidating the genetic basis of key traits, linkage mapping facilitates the development of superior mulberry cultivars, ultimately enhancing silk production and ensuring the financial viability of the sericulture industry. Leveraging a variety of molecular markers and analytical methods, linkage mapping is indispensable for comprehending the complexities of the mulberry genome and expanding breeding programs to meet the evolving needs of the silk industry. In our research, we focus on genotypes, each bringing unique benefits to our study. Thailand Male, a renowned Mulberry genotype, is favored in sericulture due to its reputation for producing high-quality leaves. Its robust leaf yield and adaptability to diverse environmental conditions offer an opportunity to explore the genetic mechanisms underlying superior leaf production. Our research, which includes Thailand Male, aims to identify genetic variations like SNPs and SSRs associated with high leaf yield. These markers could be instrumental in selective breeding programs to develop superior Mulberry cultivars that consistently provide nutrient-rich leaves for silkworms.

Assam Bol, originating from the Assam region of India, is intriguing for its inherent resistance to root rot diseases. Root rot infections pose a significant challenge in Mulberry production, often leading to reduced plant health and productivity. By studying Assam Bol, we can investigate the genetic factors contributing to its resistance to root rot pathogens. This exploration may unveil genetic, and biomarkers linked to root rot resistance, paving the way for breeding initiatives aimed at developing disease-resistant Mulberry cultivars. Such research holds practical implications, potentially reducing the reliance on chemical fungicides and promoting environmentally sustainable Mulberry agriculture.

S1, known for its exceptional drought resistance, is another crucial genotype in our study. Understanding the genetic basis of drought resistance in Mulberry is vital in the face of increasing drought conditions due to climate change. Our investigation into S1 aims to uncover genetic variations, such as SNPs and SSRs, associated with its drought resistance. This knowledge could aid in the development of drought-resistant Mulberry cultivars, ensuring a consistent leaf supply even in water-limited areas. This study not only supports the sericulture sector but also aligns with sustainable farming practices amidst changing climatic patterns.

Punjab Local, tailored to the regional characteristics of the Punjab area, plays a significant role in researching nitrogen use efficiency in Mulberry. Nitrogen is crucial for plant development, and efficient nitrogen use is essential for enhancing agricultural productivity while mitigating negative environmental impacts. By analyzing Punjab Local, we seek to identify genetic and biomarkers for effective nitrogen utilization in Mulberry. This understanding may inform strategies for optimal nitrogen management, minimizing resource wastage and environmental degradation. Furthermore, it could promote ethical agricultural practices and enhance the economic feasibility of Mulberry farming.

The inclusion of multiple Mulberry genotypes in our study broadens the scope of investigation and enables a comprehensive genetic examination of traits vital to the sericulture sector and agriculture at large. We aim to elucidate the genetic foundations of these traits using advanced methods such as SNP and SSR markers, genetic linkage mapping, and sophisticated molecular tools like DNA microarray chips. These tools also hold potential as diagnostic aids for identifying Mulberry crop issues, significantly enhancing the adaptability, productivity, and sustainability of the sericulture industry.

Our selection of four Mulberry genotypes represents a comprehensive and strategic approach to our study. The findings from our research could have implications for sustainable farming practices and the sericulture industry, particularly in the face of changing environmental conditions. By uncovering genetic markers, deciphering genetic linkage maps, and developing diagnostic tools, our study has the potential to revolutionize Mulberry agriculture, enhancing its economic viability and environmental stewardship.

CHAPTER 2

REVIEW OF LITERATURE

2.1 POLYPLOIDY IN PLANTS

Polyploidy refers to the heritable condition in which an organism possesses more than two complete sets of chromosomes. This phenomenon is observed in various organisms, including plants, certain species of fish, and amphibians [105]. The elevated chromosomal count observed in these plants is attributed to a genome duplication event. There are two primary types of polyploids distinguished by the origin of this duplication event: autopolyploid, which arise from the multiplication of a diploid genome within a single species, and allopolyploids, which occur when two distinct haploid genomes are combined through hybridization and subsequently doubled [106].

Despite the differentiation between autopolyploids and allopolyploids, the process of doubling genetic material generates a genomic buffering effect that provides advantages to both types of polyploids. This buffering mechanism safeguards the organism against the detrimental impacts of epigenetic alterations and changes in DNA sequences, thereby ensuring minimal or no adverse effects on its viability. Furthermore, this process promotes genetic diversity and facilitates the evolution of the genome [95,96]. Plant polyploidy offers significant benefits to humans, particularly considering that many major agricultural crops worldwide are polyploid. Recent genomic investigations have revealed multiple instances of genome duplication in various eukaryotic lineages, including yeast, vertebrates, and even organisms like Arabidopsis thaliana, previously believed to have a diploid small-genome structure [107].

2. 2 EXPLORING POLYPLOID GENOMES IN PLANTS

2.2.1 METHODS OF ANALYZING GENOME

The combination of genetic mapping, molecular cytogenetics, sequence analysis, and comparative analysis has enabled the discovery of numerous new understandings regarding the evolution of ploidy across various timescales. These insights range from the origins of the plant kingdom to events such as intra- and interspecific hybridization linked to plant domestication and breeding. In-depth examination of chromosomes through techniques like in situ hybridization, along with comprehensive molecular marker analyses spanning entire genomes, has yielded compelling evidence regarding the mechanisms underlying genomic modification.

The integration of genetic mapping, molecular cytogenetics, sequencing, and comparative analysis has provided valuable insights and expanded our understanding of ploidy evolution across various temporal scales. This encompasses investigations spanning the plant kingdom's foundational stages to the

occurrences of intra- and interspecific hybridization events linked to both plant domestication and breeding processes. A robust understanding of the mechanisms driving genomic modifications has been substantially enriched through the application of in situ hybridization techniques for physical chromosome analysis and comprehensive genome-wide molecular marker analyses [108].

2.2.2 *In Situ* Hybridization

It proves notably proficient in precisely discerning chromosomes and facilitating the mapping of repetitive DNA sequences as well as unique sequences onto the chromosome(s). Through the utilization of DNA probes tagged with fluorescent labels, fluorescent in situ hybridization (FISH) is employed, allowing visualization via a fluorescence microscope. Genomic in situ hybridization (GISH) employs the complete genomic DNA of a species as a probe on chromosomes, enabling discrimination of the entire genome rather than specific sequences. Investigation into the distribution of four tandem repeats in allotetraploids, such as Tragopogon mirus, Tragopogon miscellas, and their diploid progenitors, revealed evidence suggesting that chromosomal rearrangements did not occur after polyploidization, highlighting additive patterns in polyploids [109].

2.2.3 Molecular Marker-Based Genetic Mapping

Polyploid genetic mapping presents unique challenges in contrast to diploid species, primarily stemming from the intricate segregations and statistical methodologies involved. The accuracy of genetic distance measurements necessitates substantial segregating populations to address these complexities effectively [100]. Important discoveries emerged from the application of Amplified Fragment Length Polymorphism (AFLP) analysis in Arabidopsis, Brassica, and Spartina [110] or when traditional Restriction Fragment Length Polymorphism markers (RFLP) were used in B. napus, Draba norvegica, and T. miscellus, as well as wheat [111].

2.2.4 Methylation Sensitive Molecular Markers

A method akin to AFLP, employing restriction enzymes with identical recognition sites but varying sensitivity to DNA methylation (isoschizomers), has proven effective and reliable for discerning genome-wide DNA methylation patterns [110]. The technique, referred to as Methylation-Sensitive Amplified Polymorphism (MSAP), operates by targeting the 5′CCGG sequence through the isoschizomers HpaII and MspI, both sensitive to the methylation status of either the outer or inner

cytosine residues. Remarkable results were observed in newly formed polyploids through this methodology, as demonstrated by Madlung et al. in Arabidopsis [111].

2.2.5 Comparative Genome Analysis

In Brassica oleracea, around 75% of the genes underwent loss as evidenced by a comparative examination of the genomic segments. Those genes that remained were not distributed randomly; instead, they displayed dosage sensitivity. After whole-genome duplication (WGD), there was a notable over-retention of duplicates belonging to transcription factors and signal transduction pathway elements. Conversely, following smaller-scale duplications, the retention rates of these functional gene categories were comparatively reduced [112]. In four separate polyploid wheat lineages, repeated deletions of the Puroindoline (Pin) gene at the Grain Hardness (Ha) locus were identified [113]. Various methods for detecting single nucleotide polymorphisms (SNPs) have been employed to explore plant polyploidy, as demonstrated by the research [114]. The Illumina Golden Gate assay revealed a considerable abundance of single nucleotide polymorphisms (SNPs) in tetraploid and hexaploid wheat. Recently, Allen et al. further investigated this phenomenon [115]. In 6225 distinct reference sequences of hexaploid bread wheat, the Illumina GAIIx data identified upwards of 14,000 potential single nucleotide polymorphisms (SNPs). Similarly, the Illumina sequencing platform detected over one million SNPs in elite inbred lines of maize [116].

2.2.6 High-Throughput DNA Sequencing and High-Resolution Melting (HRM) Analysis

General solutions for the genetic analysis of polyploids are provided by high-throughput DNA sequencing combined with computational analysis [117]. The high level of heterozygosity made attempts to sequence the heterozygous diploid potato genome (RH89-039-16) difficult [118]. The woodland strawberry (Fragaria vesca) genome was sequenced to avoid the challenges associated with sequencing the polyploid genome of the cultivated strawberry (Fragaria x ananassa) [119].

2.3 Current strategies of Sequence alignment and SNP calling in Polyploid plants

NGS technologies, introduced in 2004, brought about a significant boost in sequencing throughput and cost reduction [120]. However, their short sequence read output also heightened the complexity of fragment assembly. Despite this challenge, NGS finds wide-ranging applications in various fields such as genome sequencing and resequencing, metagenomics, RNA-sequencing (transcriptomics), and personalized genomics (personal medicine). Notably, it allows genome sequencing even with lower

DNA concentrations. Recent NGS platforms include Roche's 454 or pyrosequencing (with read lengths up to 700 bp), Life Technologies SOLiD (50 bp), Illumina's HiSeq (two by 250 bp), and MiSeq (two by 300 bp), and Life Technologies' Ion Torrent/Proton (200 bp). Unlike Sanger sequencing, NGS methods eliminate the need for DNA cloning, simplifying the process and enhancing adaptability to various biological phenomena. This enables massive parallelization at reduced costs. However, NGS presents drawbacks such as the requirement for special assembly algorithms due to short sequence length, less accurate base calling compared to Sanger sequencing, and lower-quality assemblies than those generated from Sanger sequencing [121].

The genomes of polyploid plant species, including *Gossypium hirsutum* (cotton) and the hexaploid *Triticum aestivum* (wheat), have been sequenced using Illumina technology [122,123].

2.4 Third Generation Sequencing

Third-generation sequencing technologies, characterized by long reads, represent the latest advancements in genome sequencing. These techniques yield longer read lengths, typically measured in kilobases (Kb) rather than bases (bp), while also reducing sequencing costs and simplifying preparation and sequencing protocols [124]. This emerging technology offers numerous advantages, but it also presents drawbacks like high error rates and the requirement for exceptionally high-quality DNA samples. Nevertheless, by generating a substantial quantity of long reads to aid in resolving challenging regions within the genome, these methods show promise in addressing the hurdles associated with sequencing and assembling large, repetitive, and intricate plant genomes. Within third-generation sequencing approaches, long-read sequencing and long-range scaffolding technologies are the two main categories of techniques utilized [125].

2.5Best practices in the identification of molecular markers

The identification of molecular markers is a crucial step in various genetic studies, and following best practices ensures the reliability and accuracy of these markers. Here are key points:

2.5.1 Define Clear Objectives

Before initiating the process of marker identification, researchers must establish clear objectives for their study. A comprehensive understanding of the traits or genetic variations being targeted facilitates the selection of suitable molecular markers.

2.5.2 Choose Appropriate Marker Types

Various studies may necessitate diverse molecular markers, such as Single Nucleotide Polymorphisms

(SNPs) or Simple Sequence Repeats (SSRs), contingent upon factors like the organism under investigation, the desired extent of genetic variation, and the accessibility of genomic data.

2.5.3 Utilize High-Quality Genomic Data

The efficacy of marker identification hinges on the caliber of genomic data. Researchers ought to employ high-throughput sequencing techniques or other sophisticated methodologies to produce precise and thorough genetic data.

2.5.4 Consider Genetic Diversity

Choosing polymorphic markers is vital for capturing population diversity, enabling the differentiation between individuals or groups and offering valuable insights for genetic inquiries.

2.5.5 Validation of Markers

Identified markers should undergo rigorous validation to ensure their accuracy and reproducibility. This involves testing the markers across diverse individuals or populations to confirm their association with the target traits.

2.5.6 Evaluate Marker Transferability

In cases where markers are intended for use across different populations or species, assessing their transferability becomes crucial. Markers that can be reliably applied in diverse genetic backgrounds are highly valuable.

2.5.7 Quality Control Measures

Implementing quality control measures during marker identification is essential. This includes filtering out low-quality data, ensuring consistency in experimental conditions, and applying stringent criteria for marker selection.

2.5.8 Consider Genomic Context

Molecular markers do not act in isolation; they are part of the larger genomic context. Researchers should consider the genomic environment to understand the potential effects of neighboring genes and regulatory elements on marker functionality.

2.5.9 Document Methodology and Parameters

Transparency in methodology is vital for reproducibility. Researchers should thoroughly document the methods used for marker identification, along with any specific parameters or thresholds applied during the process.

2.5.10 Community Standards and Resources

Adhering to community standards and utilizing available genetic databases or resources can enhance the robustness of marker identification. Collaboration with other researchers and leveraging shared knowledge contributes to the establishment of best practices.

2.5.11 Ethical Considerations

Researchers should adhere to ethical guidelines in marker identification, ensuring that studies are conducted responsibly and with consideration for potential societal implications.

In conclusion, the best practices in the identification of molecular markers involve setting clear objectives, selecting appropriate marker types, utilizing high-quality genomic data, considering genetic diversity, validating markers, evaluating transferability, implementing quality control measures, considering the genomic context, documenting methodologies, adhering to community standards, and addressing ethical considerations. Following these practices enhances the credibility and utility of identified molecular markers in various genetic studies.

2.6SNP Genotyping chip for genetic studies and Breeding applications

In the realm of genetic studies and breeding applications, SNP genotyping chips have emerged as invaluable tools for understanding genetic variations and facilitating targeted breeding efforts. These chips are designed to analyze Single Nucleotide Polymorphisms (SNPs), which are subtle genetic differences occurring at a single nucleotide level. The adoption of SNP genotyping chips enables researchers to efficiently identify and study these variations across the genome [126].

The primary advantage of SNP genotyping chips lies in their ability to simultaneously assess many SNPs, providing a comprehensive view of the genetic landscape. This technology is particularly useful in genetic studies where researchers aim to identify associations between specific genetic markers and traits of interest, ranging from disease susceptibility to agricultural characteristics.

For breeding applications, SNP genotyping chips play a pivotal role in developing improved plant

varieties or animal breeds. By identifying SNPs associated with desirable traits, breeders can selectively choose individuals with the desired genetic makeup, enhancing the efficiency of the breeding process [127]. This targeted approach allows for the development of crops with increased yield, resistance to diseases, or improved nutritional content, contributing to advancements in agriculture. The success of genetic studies and SNP genotyping chip applications is intricately linked to breeding techniques and agricultural practices. Different agricultural practices can significantly influence the genetic makeup of plants, including crops like mulberry. For example, cultivation methods, soil management, and irrigation practices can impact the genetic diversity within a population. These factors may contribute to the selection of specific genetic traits over time, influencing the distribution of SNPs and markers. Understanding the agricultural practices employed in mulberry cultivation is essential for interpreting the genetic data obtained from SNP genotyping chips accurately. Selective breeding techniques also play a crucial role in shaping the genetic composition of cultivated plants. If specific breeding practices have been employed to enhance certain traits in mulberry, such as disease resistance or silk production, the associated genetic markers may be detectable through SNP genotyping [128]. Conversely, traditional breeding methods, which rely on natural processes, may result in a more diverse genetic landscape. Furthermore, the impact of environmental factors, such as climate and geographic location, on mulberry cultivation cannot be overlooked. These factors may contribute to the adaptation of certain genetic variations, potentially influencing the distribution of SNPs. In conclusion, the success of genetic studies and breeding applications, particularly those involving SNP genotyping chips, is closely tied to a comprehensive understanding of breeding techniques and agricultural practices. The interaction between these factors shapes the genetic makeup of cultivated plants, influencing the presence and distribution of SNPs and markers. A holistic approach that considers both genetic data and the context of cultivation practices is essential for leveraging the full potential of SNP genotyping chips in advancing genetic studies and breeding programs [129].

2.7 Prerequisites for the genotyping chip

In the process of developing a genotyping chip, researchers carefully select only a limited number of Single Nucleotide Polymorphisms (SNPs) to serve as markers. The choice of these specific SNPs is not arbitrary; rather, it is guided by stringent criteria. The selection process involves developing consensus sequences, essentially creating a simulated genome sequence that closely mimics the actual genome. From these mimicked sequences, confidential SNPs are derived. The criteria for choosing SNPs for the genotyping chip include quality control (QC) measures and *in silico* criteria. Quality control is a vital step in ensuring that the selected SNPs meet certain standards. Researchers set threshold values to determine which SNPs cut. This threshold value acts as a filter, allowing only those SNPs that surpass

a specified quality benchmark to be included in the genotyping chip [130]. This rigorous QC process serves as a safeguard against errors and inaccuracies, ensuring the reliability of the selected SNPs for downstream analyses. The relationship between the threshold values for SNP selection and polyploidy levels is significant, especially in the case of highly polyploid organisms like the mulberry. Polyploidy, the condition of having multiple sets of chromosomes, can complicate genetic analyses. By implementing stringent QC criteria, researchers indirectly address polyploidy issues. The QC measures help in filtering out noise and inaccuracies that may arise due to the complexity of polyploid genomes, ensuring that only high-quality SNPs are chosen for the genotyping chip. In the context of VCF (Variant Call Format) files, which store genetic variations data, there are QC and Pass/Fail columns. If a SNP falls below the predefined threshold during QC analysis, it is marked as "Fail," indicating that it did not meet the quality standards and should be excluded from further analyses. This additional layer of QC, integrated into the genotyping process, enhances the reliability of the selected SNPs and contributes to the overall success of the genotyping chip development, particularly in the context of addressing challenges associated with polyploidy, as observed in the case of the highly polyploid mulberry [131].

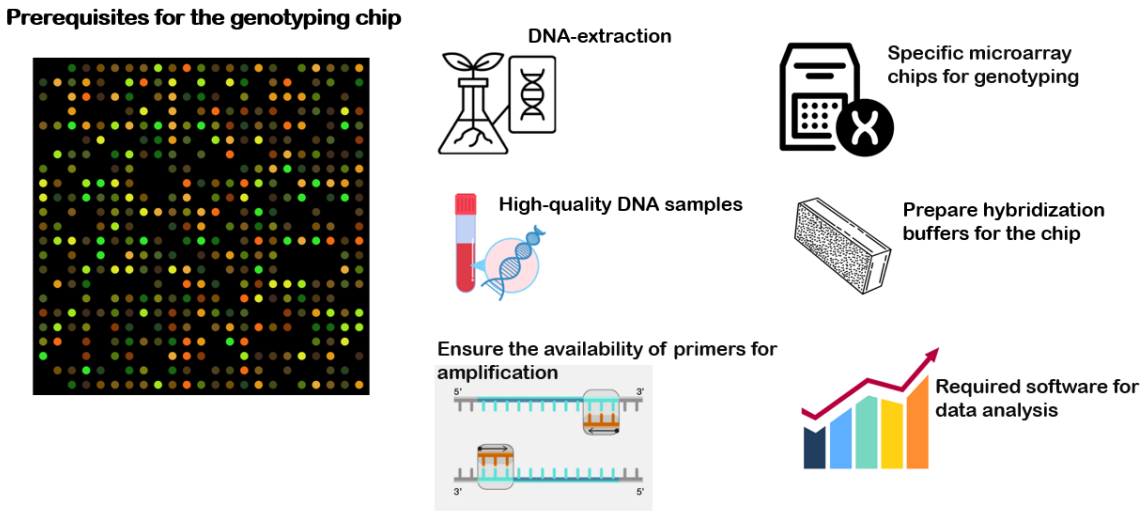


Figure 6: The picture depicts the requirements for designing a genotyping chip

2.8Linkage mapping studies

In the exploration of linkage mapping studies, researchers delved into the intricate relationships between genetic variations and observable traits. To unravel these connections, they considered four distinct genotypes and phenotypes within their samples. Each genotype represents a specific genetic makeup, while phenotypes encompass the observable characteristics or traits exhibited by the samples. The

inclusion of gene ontology, which categorizes genes based on their functions and roles, provided additional insights into the biological significance of the genetic variations under investigation. With a wealth of Single Nucleotide Polymorphism (SNP) data derived from consensus sequences, the researchers embarked on the task of linking this genetic information to the observed traits, particularly focusing on stress tolerance tests. These tests aimed to identify specific SNPs associated with stress tolerance, shedding light on the genetic basis of resilience in the studied samples. The process of connecting SNPs to gene ontology and linking them to traits involved a meticulous examination of the genetic landscape. Researchers sought to understand how variations in the SNP data correlated with the functions and roles of specific genes as defined by gene ontology. This comprehensive approach allowed for the identification of genetic markers that played a crucial role in stress tolerance, offering valuable insights into the molecular mechanisms governing this trait. Subsequently, the researchers conducted linkage mapping studies to further elucidate the connections between the identified SNPs and stress tolerance traits. Linkage mapping involves creating a map that illustrates the physical locations of genes on chromosomes and their associations with specific traits. By integrating SNP data with the results of stress tolerance tests and considering gene ontology, the linkage mapping studies aimed to provide a holistic view of the genetic landscape associated with stress tolerance in the studied samples. This integrated approach not only enhances our understanding of the genetic factors contributing to stress tolerance but also lays the groundwork for more targeted and effective strategies in breeding programs and molecular studies [132].

In traditional linkage mapping, the goal is to identify the location of genes on chromosomes. However, in this study on mulberry, the challenge emerged due to the absence of well-defined chromosomes for mulberry plants. Unlike some organisms where chromosomes are clearly identified, mulberry lacks such definite chromosomal information. Consequently, the usual approach of mapping loci on chromosomes was not feasible in this study. The absence of chromosomal mapping in the research highlights a gap in our understanding of the mulberry genome. Despite the efforts to connect SNPs with gene ontology and conduct linkage mapping studies, the lack of chromosomal information limits the comprehensive mapping of genetic loci [133].

Looking ahead, future studies could explore various avenues to address the absence of chromosomal mapping in mulberry. One approach involves further genomic research to uncover the structure and organization of mulberry chromosomes. Advances in technology and methodologies for genome sequencing may contribute to revealing the chromosomal landscape of mulberry plants. Additionally, collaborations with experts in genomics and bioinformatics could provide valuable insights into techniques that may aid in resolving the challenge of chromosomal mapping in mulberry. While the current study faced limitations in achieving chromosomal mapping, future research endeavors hold the

potential to enhance our understanding of mulberry genetics. The pursuit of novel methodologies and the integration of cutting-edge genomic tools may pave the way for comprehensive chromosomal mapping studies in mulberry plants, thereby advancing our knowledge of their genetic architecture [134].

2.9Plant Linkage mapping with genetic markers

Plant linkage mapping with genetic markers is a crucial aspect of understanding the inheritance of traits in plants. In this process, researchers analyse the association between genetic markers and specific traits, unravelling the genetic basis of various characteristics in plants. Genetic markers, such as Single Nucleotide Polymorphisms (SNPs) or Simple Sequence Repeats (SSRs), serve as signposts along the plant genome. These markers act like landmarks, helping researchers identify the location of genes responsible for traits. Linkage mapping essentially involves tracking the inheritance patterns of these markers within plant populations [135].

One significant step in plant linkage mapping is the selection of appropriate genetic markers. Researchers carefully choose markers that are distributed across the genome and exhibit variability within the plant population. This diversity ensures that the markers can effectively highlight genetic differences related to specific traits of interest. Once the genetic markers are chosen, they are applied to plant populations with known traits. The goal is to observe how the markers segregate (get passed on) along with the traits from one generation to the next. By studying the co-inheritance patterns, researchers can map the locations of genes associated with plant characteristics. In the context of plant breeding programs, linkage mapping is invaluable. It helps plant breeders identify regions of the genome that are linked to desirable traits, allowing for the targeted development of new plant varieties with improved features. For example, if a genetic marker is consistently inherited with a trait like disease resistance, breeders can selectively choose plants with that marker to enhance resistance in subsequent generations [136].

In some cases, linkage mapping might be challenging due to factors like the complexity of plant genomes or the presence of genetic variations. Advanced technologies, such as Next-Generation Sequencing (NGS), have significantly enhanced the precision and efficiency of linkage mapping studies, enabling researchers to delve deeper into the genetic architecture of plants. In conclusion, plant linkage mapping with genetic markers is a powerful tool in deciphering the genetic basis of plant traits. It contributes to our understanding of inheritance patterns, aids in targeted plant breeding, and facilitates the development of improved crop varieties with desirable characteristics. The integration of cutting-edge technologies further strengthens the accuracy and scope of plant linkage mapping studies, opening

new avenues for advancements in plant genetics and breeding [137].

2.10 Computational requirements for linkage mapping studies

Computational requirements play a crucial role in facilitating linkage mapping studies, especially in the context of unravelling the genetic basis of traits in various organisms. These studies involve the analysis of large datasets, complex statistical computations, and intricate bioinformatics analyses, demanding significant computational resources. One fundamental aspect of computational requirements in linkage mapping studies is the handling of massive genomic datasets. As researchers utilize advanced technologies like Next-Generation Sequencing (NGS), they generate vast amounts of genetic data. Efficient storage solutions and powerful computing infrastructure are essential for managing and processing these large datasets. High-performance computing clusters or cloud-based platforms often become indispensable for handling the computational load associated with genomic data analysis [138]. The algorithms employed for linkage mapping are computationally intensive, requiring substantial processing power. These algorithms assess the genetic markers' segregation patterns and calculate statistical probabilities to identify regions of the genome associated with specific traits. Parallel computing capabilities, which allow multiple calculations to be performed simultaneously, can significantly accelerate the execution of these algorithms. Additionally, bioinformatics tools are integral to extracting meaningful insights from genomic data. Variant calling, genotype imputation, and quality control procedures involve complex computational procedures. Adequate computational resources are necessary to execute these bioinformatics analyses accurately. Moreover, the visualization of linkage maps and the interpretation of results often involve software applications that demand substantial computational capabilities. The size and complexity of the genome being studied also impact computational requirements. For organisms with large and intricate genomes, such as some plants, animals, or humans, the computational demands escalate. As researchers aim for higher resolution and accuracy in mapping genetic loci, the computational infrastructure must be robust enough to handle the intricacies of such genomes. As technologies evolve, the computational demands of linkage mapping studies continue to grow. The integration of machine learning approaches, advanced statistical methods, and sophisticated algorithms further accentuates the need for powerful computational resources. Collaboration with computational experts, access to supercomputing facilities, and utilization of cloud computing services are common strategies employed by researchers to meet these computational requirements. In summary, computational requirements for linkage mapping studies encompass the management of large genomic datasets, execution of complex algorithms, utilization of bioinformatics tools, and adaptation to the size and complexity of the studied genome. Adequate computational resources are essential for the successful execution and interpretation of linkage mapping studies,

contributing significantly to advancements in genetics and genomics research [139].

2.11 MOTIVATION OF THE STUDY

The motivation behind this research stems from a deep-seated commitment to revolutionize mulberry cultivation practices, particularly in the context of Indian varieties of *Morus indica*. The current agricultural landscape necessitates a multidimensional exploration into the genomic makeup of these varieties to enhance their cultivation methodologies effectively [140,141].

Our endeavor is driven by the pressing need to address key challenges faced by mulberry growers, including resilience to drought conditions, optimization of leaf yield, enhancement of nitrogen efficiency, and resistance to root rot. These challenges significantly impact agricultural productivity and sustainability, underscoring the importance of genetic insights in informing targeted breeding efforts and selection strategies.

Furthermore, the lack of a comprehensive reference genome or datasets for these unique *Morus indica* varieties presents a significant gap in knowledge. This research seeks to bridge that gap by employing cutting-edge techniques such as genotyping, consensus generation, linkage mapping, SNP generation, SSR markers, and microarray chip development. By pioneering the utilization of a draft reference genome for computational analyses, we aim to unlock vital genetic information that can drive transformative advancements in mulberry cultivation.

At the heart of this research lies a commitment to innovation and progress. The development of a specialized gene chip tailored to mulberry cultivation represents a groundbreaking advancement, promising to provide growers with actionable insights to optimize their practices and enhance crop yields. Moreover, the establishment of an expansive genetic database will serve as a cornerstone for future research endeavors, fostering continued exploration into the genetic intricacies of *Morus indica* [142].

Despite the formidable challenges posed by the scarcity of genetic data and the absence of a reference genome, our dedication remains unwavering. We believe that the anticipated findings of this research have the potential to catalyze a paradigm shift in mulberry cultivation practices, not only within India but also on a global scale. By fostering sustainability and resilience in sericulture industries worldwide, we aim to make a meaningful and lasting impact on agricultural landscapes.

In essence, this research is driven by a passion for innovation, a commitment to addressing pressing agricultural challenges, and a vision for a more sustainable and resilient future for mulberry cultivation.

2.12 SCOPE OF THE STUDY

The proposed research aims to pioneer a multidimensional exploration into the genomic landscape of Indian varieties of *Morus indica*, with a particular emphasis on enhancing mulberry cultivation practices. By employing cutting-edge techniques such as genotyping, consensus generation, linkage mapping, SNP generations, SSR markers, and microarray chip development, the research endeavors to unravel the genetic intricacies underlying four distinct phenotypes: S1, Thailand Male, Punjab Local, and Assama Bola. Additionally, the investigation will delve into four specific genotypes characterized by traits crucial for agricultural productivity: resilience to drought conditions, optimization of leaf yield, enhancement of nitrogen efficiency, and resistance to root rot. Given the dearth of a reference genome or comprehensive datasets for these unique *Morus indica* varieties, the research represents a pioneering effort in utilizing a draft reference genome to undertake sophisticated computational analyses. The overarching objective is to harness genomic insights to engineer a specialized gene chip, tailored to augment mulberry cultivation practices and stimulate agricultural growth. The research objectives are multifaceted and encompass various critical facets. Firstly, the endeavor seeks to unravel the intricate genetic makeup of the target phenotypes, aiming to discern common genetic elements through consensus sequence generation. Subsequently, the construction of detailed genetic maps will elucidate the intricate relationship between key traits and molecular markers, including the identification of Single Nucleotide Polymorphisms (SNPs) and Simple Sequence Repeats (SSRs) [143,144]. This phase holds the potential to unveil crucial insights into the genetic architecture governing mulberry traits, facilitating targeted breeding efforts and informed selection strategies. A pivotal aspect of the research involves the development of a bespoke microarray chip, engineered to enable rapid and comprehensive gene expression profiling. This transformative technology promises to revolutionize mulberry cultivation practices by providing growers with actionable insights to optimize cultivation methodologies and enhance crop yields. Furthermore, the research aims to establish an expansive genetic database, integrating genetic mapping data, microarray chip results, analysis outcomes. This comprehensive repository will serve as a cornerstone for ongoing research endeavors, providing a robust foundation for further exploration into the genetic intricacies of *Morus indica*. Despite the formidable challenges posed by the scarcity of a draft reference genome and the limited existing data on these specific *Morus indica* varieties, the research remains steadfast in its commitment to making substantial contributions to the burgeoning field of genomics and mulberry cultivation. The anticipated findings hold the promise of catalyzing a paradigm shift in mulberry cultivation practices, not only within India but also across global agricultural landscapes, fostering sustainability and resilience in

sericulture industries worldwide [145-148].

2.13 RESEARCH GAP

Despite the importance of *Morus indica* in agriculture and sericulture, there is a notable absence of a centralized database containing the complete genome sequence of Indian varieties. Existing genomic resources primarily focus on non-Indian mulberry species, which may not fully capture the genetic diversity present in Indian populations. The availability of molecular markers tailored to Indian mulberry varieties is limited. Existing marker databases predominantly comprise markers designed for non-Indian genetic backgrounds, which may not accurately reflect the genetic architecture of Indian *Morus indica* populations [149,150]. As a result, the utility of these markers for assessing genetic diversity and guiding crop improvement efforts in Indian contexts is compromised. Despite advancements in genomic research, there is a significant gap in translating research findings into actionable insights for farmers. Farmers lack access to information regarding the traits of interest and strategies for crop improvement based on genomic insights [151]. The absence of user-friendly tools, such as a chip for assessing genetic diversity, further hinders farmer engagement and adoption of genomic technologies in mulberry cultivation. Bridging the gap in genomic resources and molecular marker information for *Morus indica* is crucial for accelerating crop improvement initiatives and enhancing agricultural productivity in India. Establishing an Indian-specific database will not only advance our understanding of the genetic basis of desirable traits but also empower farmers with actionable insights for sustainable mulberry cultivation. This research endeavor represents a vital step towards harnessing the full potential of genomic technologies to benefit Indian agriculture and sericulture [152,153].

OBJECTIVES OF THE STUDY

Based on the Research gap and the Scope of the study, 3 Main objectives were framed for the Research study.

1) Development of database with the curated SNP and SSR Markers
2) Construction of the linkage maps with the Markers
3) Development of SNP Genotyping Chip

CHAPTER 3

RESEARCH METHODOLOGY

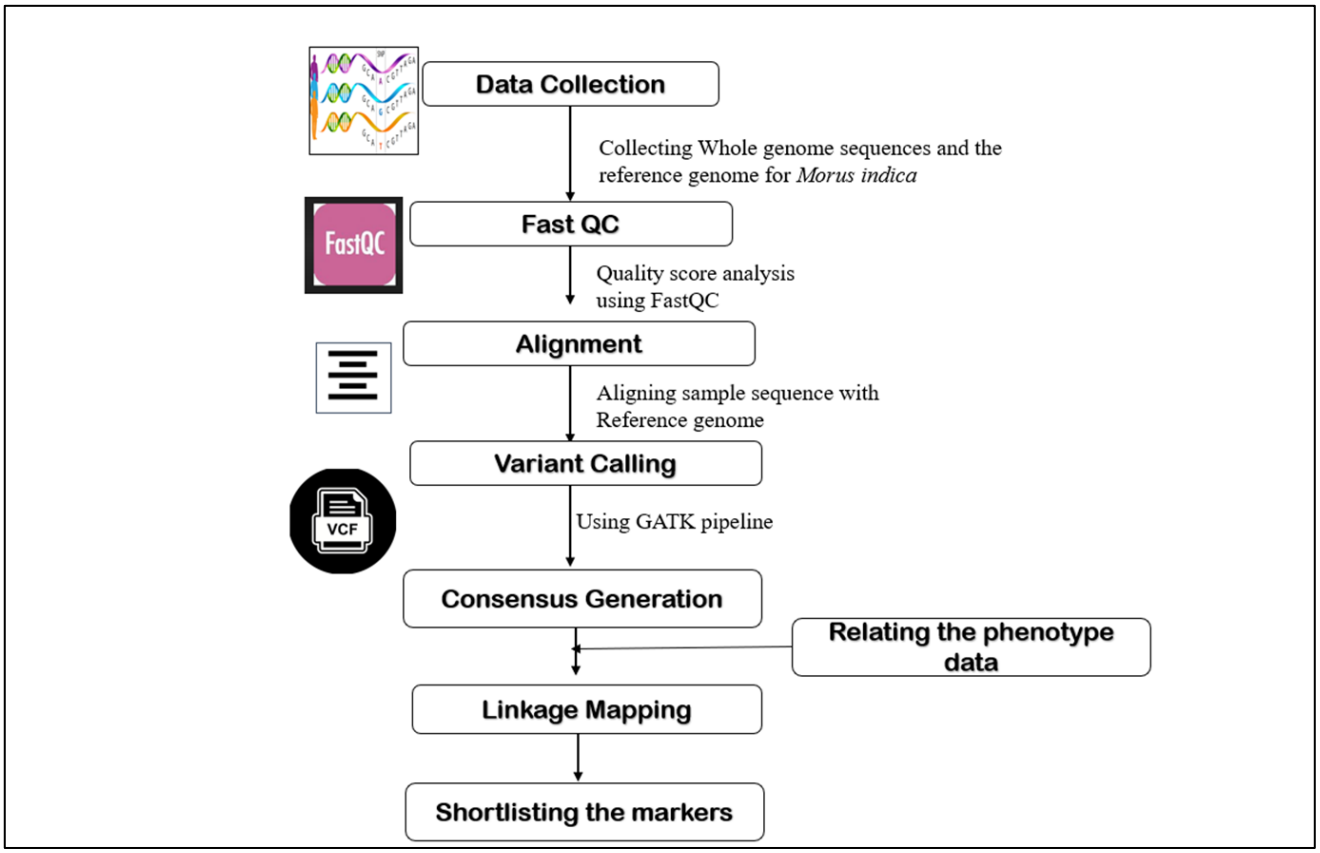


Figure 7: Overview of the Methodology followed

OBJECTIVE 1: Development of database with the curated SNP and SSR Markers

3.1 Whole genome sequencing of the Mulberry data

Whole genome sequencing (WGS) with Illumina technology offers a powerful approach for exploring the genetic makeup of Morus indica, the mulberry tree. This species holds immense value in silk production and traditional medicine. Illumina's high-throughput sequencing provides extensive genomic coverage rapidly and affordably. Its accuracy ensures precise reconstruction of the genome, crucial for identifying genetic variations like SNPs. Moreover, the scalability of Illumina sequencing allows researchers to tailor their study objectives, whether uncovering genetic markers or tracing evolutionary paths. Ultimately, this method enhances our understanding of *Morus indica's* genetic landscape, unlocking opportunities for agriculture and medicine.

The whole genome sequencing of the Mulberry was carried out using Illumina sequencer. Total of 16 genotypes were sequence and 4 were selected for the analysis due to their correlation with the phenotypes selected i.e, Root rot, Nitrogen use efficiency, Drought and Leaf yield. The details of which are shared in the table 2.

K2 is renowned for its high yield and adaptability to diverse climates, stemming from a cross between local Indian rice varieties. V1 stands out for its disease resistance, particularly against brown spot and sheath blight, appealing to farmers seeking robust yields. S13 excels in drought resistance, crucial for cultivation in arid regions. AR12 showcases resistance against pests like the rice borer, ensuring crop protection. Similarly, C176, S34, S1635, and S1 offer disease and pest resistance, enhancing yield reliability. RFS175 further fortifies against pests, particularly the rice borer. *M. multicaulis*, or Mung bean, enriches soil fertility through nitrogen fixation, boosting overall crop productivity. Himachal Local, native to Himachal Pradesh, India, exhibits resilience against diseases and pests. BR-8, like other varieties, combines disease resistance with high yield potential. These varieties collectively empower farmers with options tailored to environmental challenges, ensuring sustainable agricultural practices and food security [154]. The correlation of the genotype and phenotype are mentioned in the table 1.

Table 1 The Genotype details

Genotype name	Drought	Leaf Yield	NUE	Root rot
K2	Medium	Low	-	Moderately Resistant
V1	High	High	High	Highly susceptible
S13	Medium	Medium	Moderate	Resistant
AR12	Not included	Medium to high	-	Resistant
C176	Not included	High	-	Moderately Resistant
S34	Not included	Medium	High	Resistant

S1635	Not included	High		Resistant
S1	Not included	Medium	High	Highly susceptible
RFS175	High	High	Low	Resistant
M. multicaulis	Not included		-	Highly Resistant
Himachal Local	Not included	Low	-	Resistant
BR-8	Not included	Low	-	Moderately Resistance

3.2 Retrieval of raw mulberry data

The high-depth sequencing of multiple mulberry accessions via Illumina platform was performed by Jawaharlal University. A total of 21 genotypes were sequenced and deposited in the *Morus indica* Genome project (http://tgsbl.jnu.ac.in/MindGP/). Out the 21 genotypes 4 were selected for analysis which had impact on the phenotypes i.e. Root rot, Leaf yield, Nitrogen use efficiency and Drought. The details of the phenotypes are mentioned in the table below:

Table 2 Phenotype details of the samples collected

Genotype	Drought	Leaf Yield	Nitrogen use Efficiency	Root rot
S1	Low	High	Medium	Highly Susceptible
Thailand Male	Low	Medium	Moderate	Highly Susceptible
Punjab Local	High	Medium	Nil	Highly Susceptible
Assama Bola	Medium	High	Low	Moderate

3.3 Quality checks of the reads

We wanted to make sure the raw data we gathered was up to scratch, so we ran it through a quality assessment using tools like FastQC and MultiQC. FastQC is great for giving us a quick overview of

any issues in the raw sequencing data. We checked all four genotypes of *Morus indica* to see if there were any problems like inconsistent quality across sequences or unusual nucleotide content. After running these checks, we knew which datasets were good to go for further analysis. FastQC also helped us spot any over-represented sequences, which might indicate contamination from adapters or primers. This detailed report gave us insight into whether we needed to do any additional steps like trimming bases, clipping adapters, or filtering reads before moving on to alignment.

3.4 Gapped alignment

To make sure we accurately and quickly matched up the pre-processed reads with the reference genome, we used a process called alignment mapping. There are different tools available for this task, but we chose BWA (Burrows-Wheeler Aligner) because it's known for its speed and efficiency. BWA uses a clever technique called the Burrows-Wheeler Transformation along with the Smith-Waterman dynamic programming algorithm to align the short reads with the reference genome. It has three different algorithms for this job: BWA-backtrack, BWA-MEM, and BWA-SW. By default, BWA looks for alignments within a certain distance from the query sequence, but it can also be adjusted to find longer gaps if needed, although this might slow things down a bit and increase the chance of incorrect matches. The reason we went with BWA for this task is that it's really fast. It can handle about two million high-quality short reads and map them against the mulberry genome in just around 30 minutes.

3.5 Refining the alignments and conversion by SAM tools

After aligning the cancer exome short reads to the reference genome, we wanted to enhance the quality of the alignments and minimize the chance of errors in variant calling. To achieve this, we ran the alignment output through several refining steps using tools designed for processing SAM/BAM files. There are various software options available for handling SAM/BAM files, including SAM tools, Genome Analysis Toolkit (GATK), and Picard. These tools, developed by institutes like Broad and Sanger, are commonly used for operations on alignment data. In our study, we refined the alignments for all twenty datasets. This involved tasks such as sorting the reads, recalibrating quality scores, marking PCR and optical duplicates, realigning indels, and filtering the reads. SAM tools is a versatile toolkit specifically designed for manipulating and analyzing alignment data in SAM/BAM format. It can handle tasks like converting from other alignment formats, merging and sorting alignments, removing PCR duplicates, calling variants like SNPs, and providing per-position information. We used SAM tools to convert the short-read alignment output from SAM to BAM format. Then, we sorted and merged the BAM files. SAM format is widely used due to its simplicity and flexibility, while BAM, being a binary representation, offers a more compact size and faster retrieval of alignments. By leveraging positional indexing, sorting, and merging, SAM/BAM files allow for efficient processing of genomic regions without loading the entire file into memory. This approach enables an integrated

analysis of genome sequence data, separating the alignment step from downstream analyses. Therefore, our study employed SAM tools to convert SAM to BAM format, facilitating subsequent analyses.

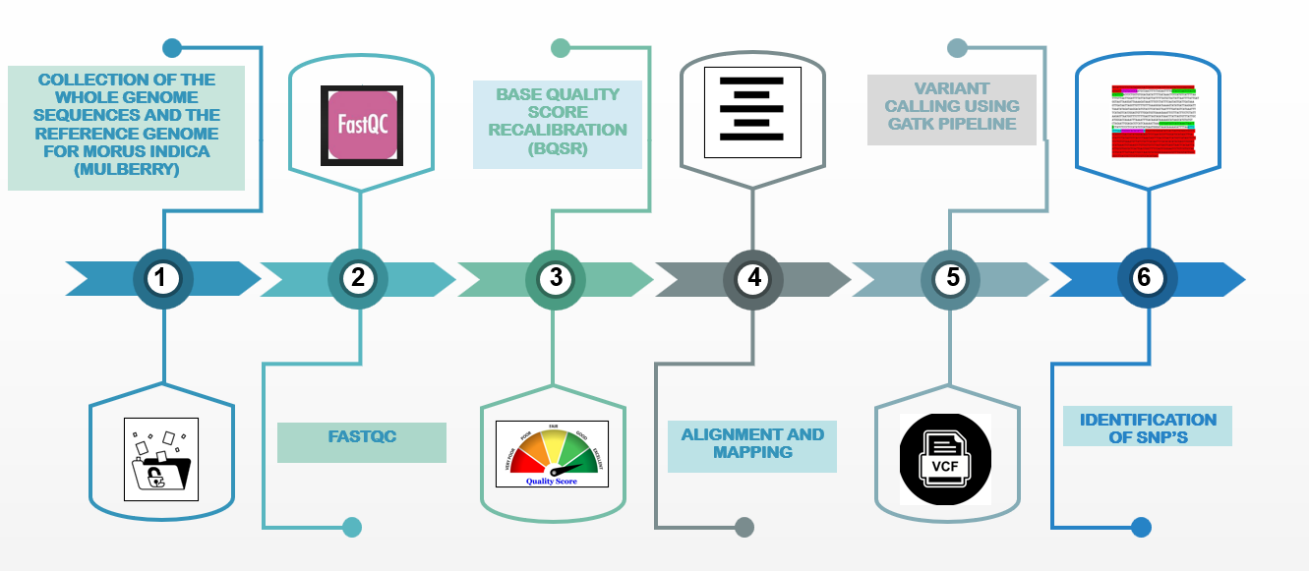


Figure 8: Protocol for Mapping and SNP calling

3.6 Variant processing

First, before identifying any mutations or variants from the sequenced data, we go through some initial steps. This includes indexing the reference genome using tools like BWA and SAM tools, as well as checking the quality of the sequences using FASTQC. After that, we merge the alignments in BAM format. Once we have these preliminary steps done, we move on to variant calling and processing. In our study, we started this process using two main tools: PICARD and GATK. GATK, short for Genome Analysis Toolkit, is a framework designed to make it easier to develop effective tools for analysing DNA from next-generation sequencing (NGS). It offers a wide range of data access patterns needed for most analyses, and its structure allows for optimization in terms of precision, memory efficiency, and stability. GATK also provides tools for calculating coverage and identifying single nucleotide polymorphisms (SNPs). In our study, we used GATK to get rid of PCR duplicates, which is essential for efficient and accurate analysis, especially in large-scale projects. On the other hand, PICARD is a set of command-line tools commonly used for handling high-throughput sequencing data and formats like SAM, BAM, CRAM, and VCF. Its file formats are well-defined in the HTS-specs repository. In our study, we used PICARD to mark PCR duplicates and perform local realignment and base quality recalibration using parameters suited for our analysis. We also used PICARD for tasks like sorting SAM files, recalibrating bases, and building BAM indexes. Once we analyzed the covariates, we proceeded to call the variants.

3.7 Variant calling

After processing the variants, we proceeded to call them. To do this, we first piled up the mapped data and counted the bases using SAMtools. Then, we used GATK to identify and extract the SNPs and indels from the variants. Once we had all the variants identified, we focused on calling the SNPs. Next, we extracted the SNPs and indels and filtered them to ensure accuracy. We then used a tool called snpEFF to annotate the filtered SNPs and indels. SnpEFF is handy because it not only annotates the variants but also predicts the effects of these genetic changes, including any alterations in amino acids. SnpEFF takes in variant call format (VCF) files containing the predicted variants, such as SNPs and indels, and outputs its annotations in the default VCF format. In our study, we carefully analyzed the identified SNPs and scrutinized the output from snpEFF. One of the advantages of using snpEFF is its speed—it can annotate up to 1,000,000 mutations per minute, which is quite impressive. Additionally, it supports over 2500 genomes, ranging from plants and animals to bacteria and fungi. Another benefit is its seamless integration with GATK and Galaxy toolkits. Overall, we found snpEFF to be a valuable tool for both variant calling and annotation in our study.

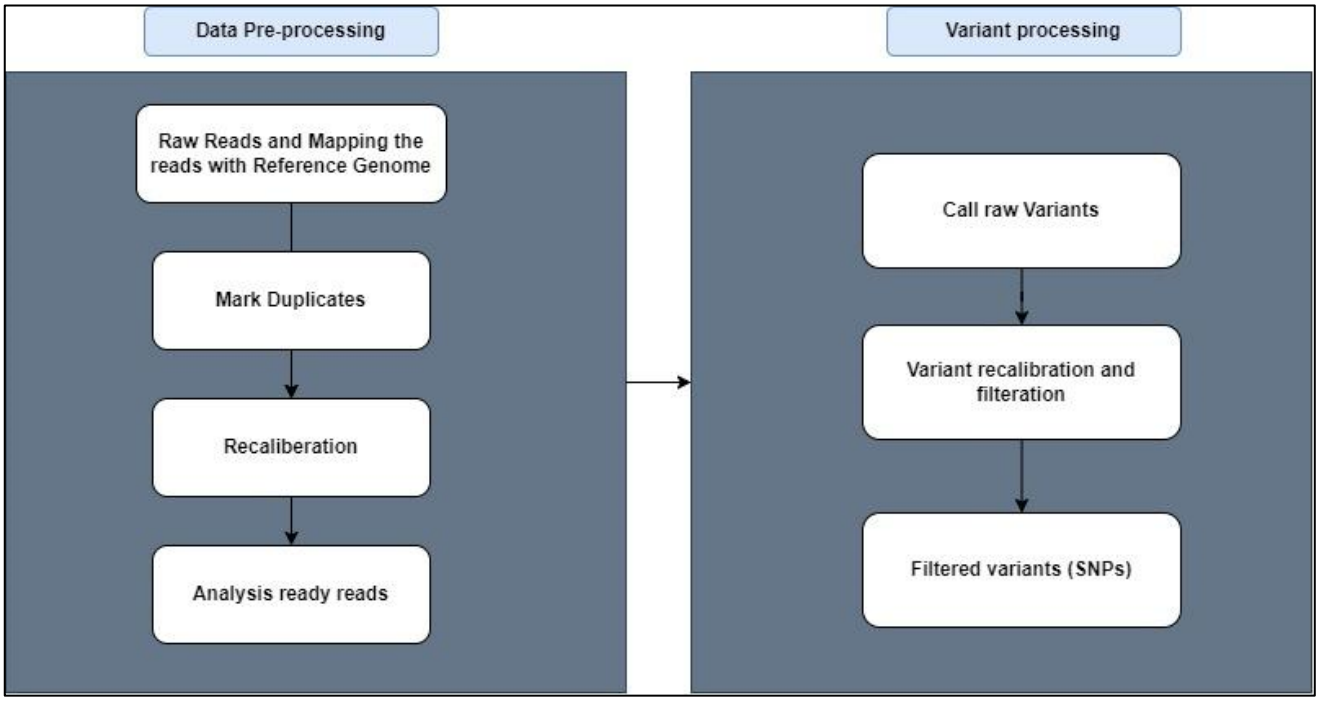


Figure 9: Best Practices for GATK pipeline for calling the variants

3.8 Variant Filtration

Understanding the link between genotype and phenotype is a key scientific challenge, and the ability to anticipate phenotypes based on molecular genotypes is critical in molecular breeding. Whole genome duplications have affected the history of all flowering plants and pose difficulties in determining the link between genotype and phenotype, particularly in polyploid species. Although single nucleotide

polymorphisms (SNPs) have become popular tools for genetic mapping, finding and using SNPs in polyploids has proven challenging. These findings are expected to differ amongst species; hence we suggest a set of best practices for SNP calling in polyploids.

Filtering criteria used in variant calling analysis, particularly for single nucleotide polymorphisms (SNPs). Each criterion targets specific aspects of variant quality and potential artifacts within the data. Variants with low confidence due to factors like poor quality by depth (QD < 2.0), high strand bias (FS > 60.0 and SOR > 4.0), low mapping quality (MQ < 40.0), and discrepancies in mapping quality between reference and alternate alleles (MQRankSum < -12.5) are filtered out. Additionally, variants where the alternate allele is found at the ends of reads more frequently than the reference allele (ReadPosRankSum < -8.0) are also removed. These filtering steps help ensure the reliability of variant calls by removing potential false positives caused by technical artifacts or sequencing errors.

3.9 Identification of SSR Markers

Genome Wide Microsatellite Analyzing Tool was used for analyzing the gene wise Markers

Data Preparation

The genomic data utilized in this study were obtained in FASTA format, providing the foundational sequences for the organism under investigation. The GMATA software was subsequently installed on the local computer system to facilitate the analysis.

Running GMATA

To initiate the SSR identification process, a terminal or command prompt was accessed on the computer. The terminal was navigated to the directory where the GMATA software was installed, ensuring that the tool was readily accessible for analysis.

Identifying SSRs

The cornerstone of SSR identification within GMATA is the Microsatellite identification tool (MISA) module. This algorithm systematically detects SSRs within the genomic sequences based on user-defined parameters. Specifically, repeat unit length and the minimum number of repeats were customized to suit the requirements of our investigation.

3.10 DATABASE DEVELOPMENT

3.10.1 Frontend Development

The frontend of the Mulberry Database project facilitates seamless retrieval of genetic data based on specific key-value pairs, enhancing user interaction and data exploration. Here's a detailed overview of the frontend architecture, functionalities, and user experience:

Architecture

3.10.2 Next.js and Tailwind CSS: The frontend leverages Next.js and Tailwind CSS to create a responsive and visually appealing user interface. Next.js enables server-side rendering for enhanced performance, while Tailwind CSS provides a utility-first approach for streamlined styling.

- **Axios for HTTPS Requests**: Frontend-to-backend communication is managed via Axios, enabling asynchronous retrieval of genetic data based on user input.
- **Data Retrieval Based on Key-Value Pairs**

The frontend enables users to retrieve genetic data by specifying key-value pairs, such as selecting the 'id' attribute and providing the value 'Mi00088'.

For instance, upon submitting the search query for the specified key-value pair, the frontend displays the following result:

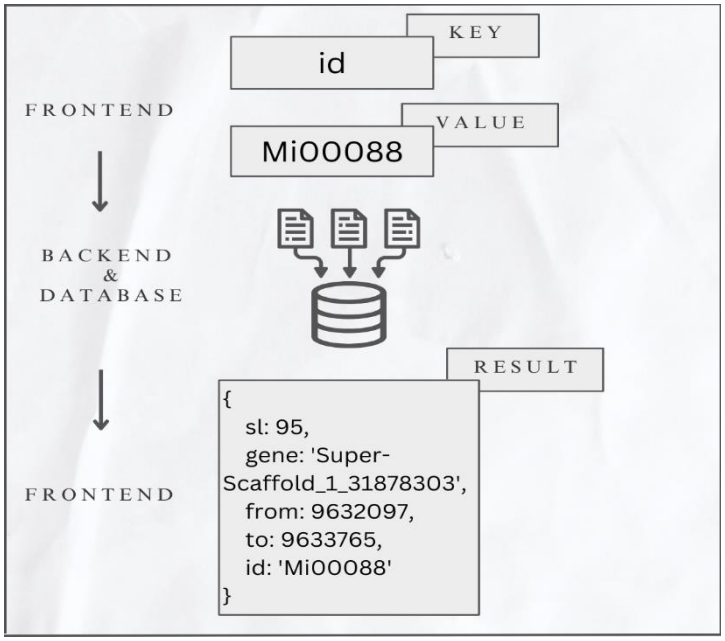


Figure 10 The query flow for fetching data (gene sequences) from database

The frontend interface restricts the display to values pertinent to the search query, ensuring that users receive focused and relevant genetic data matching their criteria.

Users can easily select attributes and input corresponding values to retrieve specific genetic sequences, enhancing the efficiency and effectiveness of data exploration and analysis.

- **Hosting Platform**

The frontend application is hosted on **Netlify**, a reliable hosting platform offering continuous deployment and scalability capabilities.

Netlify's seamless deployment process ensures that updates and enhancements to the frontend are promptly reflected in the live application environment, providing users with a consistent and up-to-date experience. The application is live on mulbery_application

3.10.3 Backend Development

The backend of the Mulberry Database project, powered by **Node js** and **Express js**, orchestrates communication between the frontend and the MySQL database hosted on **Aiven Cloud.** Here's an overview of the backend architecture, functionalities, and hosting platform:

Architecture

Node.js and Express: The backend utilizes Node js and Express js to create a robust and scalable server environment, capable of handling incoming **HTTPS** requests from the frontend.

MySQL Database Interaction: The backend interacts with the MySQL database hosted on Aiven Cloud, executing queries and managing data retrieval and insertion operations.

3.10.4 Advantages

- **Efficient Data Management**: The backend facilitates seamless data management, allowing users to upload genetic data via **CSV files** and add values to the database securely. The backend ensures data integrity and consistency, maintaining the reliability of genetic information stored in the database.
- **Security and Access Control**: Authentication mechanisms implemented in the backend enforce security measures, preventing unauthorized access and manipulation of sensitive genetic data. By implementing access controls, the backend enhances data confidentiality and privacy.

3.10.5 Hosting Platform

The backend application is deployed on **Render.com**'s free service tier, providing a reliable and scalable hosting platform for Node.js applications. Render.com's managed infrastructure simplifies deployment and maintenance tasks, ensuring high availability and performance of the backend server.

3.10.6 Database Management

The Mulberry Database project relies on MySQL hosted on Aiven Cloud to manage and store genetic data efficiently. Here's a detailed breakdown of the database architecture, features, and advantages:

3.10.7 Database Structure and Attributes

The Mulberry Database comprises a single table named MULBERRY_DB with the following attributes:

sl: An integer field serving as the primary key with auto-increment functionality.

gene: A varchar field with a maximum length of 255 characters, allowing for the storage of gene identifiers.

from1: An integer field representing the starting position of a genetic sequence.

to1: An integer field representing the ending position of a genetic sequence.

id: A varchar field with a maximum length of 255 characters, potentially storing additional identifiers.

s1: A text field for storing textual genetic data.

Assama_Bola: A text field for specific genetic attributes.

Thailand_Male: A text field for specific genetic attributes.

Punjab_Local: A text field for specific genetic attributes.

ssr_repetitions: A varchar field for storing repetitions of specific genetic sequences.

ssr_motif: A varchar field for storing motifs of specific genetic sequences.

Gene_ontology_IDs: A text field for storing gene ontology identifiers.

Molecular_Function: A text field for storing molecular function descriptions.

3.10.8 Database Size and Performance

Size: The Mulberry Database table contains approximately **26,964 records**, comprising genetic data totaling around **1.2 gigabytes**. This size demonstrates the scale and complexity of the genetic information managed by the database.

Performance: MySQL on Aiven Cloud offers robust performance and scalability, ensuring optimal query execution and data retrieval even with large datasets. With horizontal scalability and automatic resource provisioning, the database seamlessly accommodates growing data volumes and user demands without sacrificing performance.

3.10.9 Advantages of MySQL on Aiven Cloud

- **Scalability and Performance**: Aiven Cloud's MySQL service provides horizontal scalability, enabling the Mulberry Database to handle increasing data volumes and user traffic efficiently. Automatic resource provisioning ensures optimal performance and responsiveness, even under heavy loads.
- **Data Security and Compliance**: Aiven Cloud implements advanced security measures, including encryption at rest and in transit, to safeguard the confidentiality and integrity of genetic data stored in the MySQL database. Compliance with regulatory standards such as GDPR and HIPAA ensures adherence to data protection requirements and user privacy.
- **Managed Services**: Aiven Cloud's managed database services simplify database administration tasks, including monitoring, backups, and maintenance. Automated management features reduce operational overhead and ensure the reliability and availability of the Mulberry Database.

3.10.10 Hosting Platform

The MySQL database is hosted on Aiven Cloud, a leading provider of managed database services. Aiven Cloud's scalable infrastructure and comprehensive management capabilities make it an ideal platform for hosting critical databases like the Mulberry Database. With high availability, data durability, and security features, Aiven Cloud ensures the reliability and integrity of genetic data stored in the database.

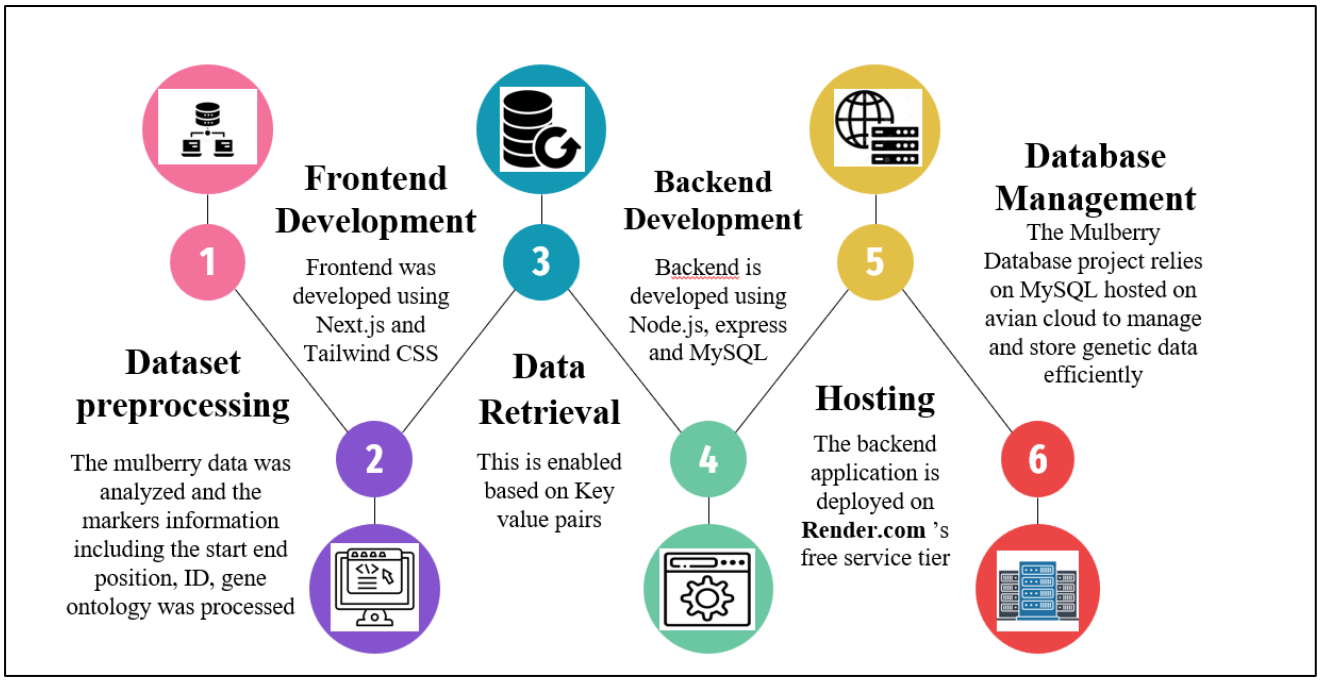


Figure 11: The flow for the Database development

Objective 2: Construction of the linkage maps with the Markers

3.11 Consensus Generation

To generate a consensus sequence based on variant calls from a VCF (Variant Call Format) file and a reference genome, several steps are typically involved. Here's a write-up outlining the methods followed for consensus generation based on the provided commands:

Variant Calling and Filtering

The first step involves variant calling from aligned sequencing data (e.g., from DNA sequencing reads). Variant calling identifies positions in the genome where individuals in the sample differ from the reference genome.

The file **vcf** likely contains variant calls, specifically Single Nucleotide Polymorphisms (SNPs), filtered to retain high-quality variants. This file may have been generated using tools such as GATK,

Compression and Indexing:

The **bgzip** command is used to compress the VCF file and give the output in the form of vcf.gz

Following compression, the **tabix** command is applied to index the compressed VCF file for efficient retrieval of variants. The index is generated in VCF format.

Reference Genome Preparation:

The mulberry reference genome is prepared for use in generating the consensus sequence. This file likely contains the entire genomic sequence of the organism under study, broken down into contigs or scaffolds.

Consensus Sequence Generation:

The **samtools faidx** command indexes the mulberry reference genome to facilitate random access to specific regions of the genome.

Following indexing, the **bcftools consensus** command is utilized to generate a consensus sequence for a specific region of interest. In this case, the command generates a consensus sequence for "Super-Scaffolds" based on the provided reference genome and variant calls from the SNP and indel VCF files (the **filtered snps in vcf.gz**)

The output of this command is redirected (>) to a FASTA file named **Contig.fa**, which contains the consensus sequence for the specified region.

In summary, the methods involve variant calling and filtering, compression and indexing of VCF files, preparation of the reference genome, and generation of a consensus sequence based on the reference genome and variant calls using specific commands from the bcftools and samtools software packages. These steps enable the creation of a consensus sequence that reflects the most likely genomic sequence for the specified region, considering both SNPs and indels detected in the sequencing data.

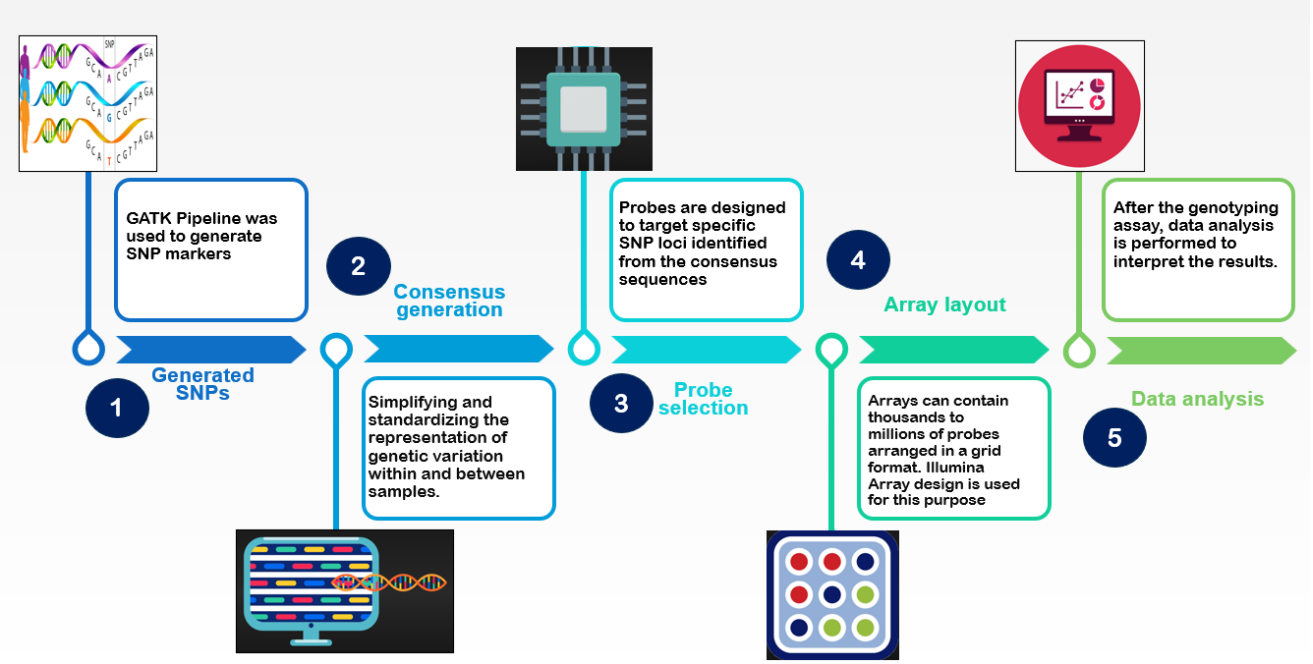


Figure 12: Workflow for Consensus generation with the SNPs

3.12 Designing the primer sequences

Once the consensus has been developed, the primers were designed for the SNPs obtained using Primer3 design suite. To obtain the right set of primer sequences, the following parameters were set:

a) Primer minimum size was set to 18 and maximum was set to 30

b) The melting temperature was set to 52

c) Primer GC content was set to 20

The parameters were set and 25 bases before and after the SNP position (considered as flanking sequences) was given as a sequence for the design of primers.

3.13 Gene ontology studies and genotype-phenotype relation

Gene Ontology (GO) analysis is an essential component of our study, facilitating a systematic exploration of gene function and biological processes associated with key phenotypic traits in *Morus indica* varieties. QuickGO, a user-friendly online tool provided by the Gene Ontology Consortium, is utilized for this purpose. The Pathways relating to the phenotypes considered in the study were considered. A correlation matrix, depicting the relation among the phenotypes was drawn.

3.14 LINKAGE GROUPS

To prepare for genetic mapping, we begin by creating a mapping population, typically through crosses between different *Morus indica* varieties. This population, often bi-parental (like F2, RILs, or DH lines), ensures genetic diversity for robust linkage mapping outcomes. It's crucial to ensure a sufficiently large population size to capture the genetic variation adequately. Next, we genotype the mapping population using molecular markers suitable for linkage mapping, such as SSRs or SNPs. These markers are chosen based on their polymorphism and genome coverage to provide comprehensive genetic information. The marker data is then formatted to be compatible with QTL IciMapping Version 4.2, the software we'll use for analysis. With the marker data in hand, we utilize QTL IciMapping Version 4.2 to construct linkage maps. This involves selecting appropriate mapping algorithms and parameters to accurately represent the genetic distances between markers. The quality of the constructed linkage maps is assessed, ensuring marker order and map distances are reliable for subsequent analysis.

Objective 3: Development of SNP Genotyping Chip

3.15 CHIP DEVELOPMENT

To begin designing an effective SNP genotyping assay, it's essential to clearly define the objectives of the assay. This involves identifying the specific SNPs of interest, understanding their genomic positions, allelic variants, and any relevant flanking sequences. Accuracy and completeness of the SNP data are crucial for precise assay design.

Once the SNP data is gathered, accessing the Design Studio Assay Design Tool becomes the next step. This tool, provided by the manufacturer or developer, offers a platform specifically tailored for SNP genotyping. Logging in and accessing the Assay Design Tool within Design Studio facilitates the customization of assay parameters according to the requirements of the study.

Assay parameters such as SNP positions, primer sequences, probe sequences (if applicable), amplicon size, and melting temperature (Tm) need to be specified. It's important to consider the genotyping platform and the specific requirements for primer and probe design. For instance, if using a PCR-based assay or microarray probes, compatibility with these platforms should be ensured for optimal assay performance.

By carefully defining objectives, preparing accurate input data, and accessing the appropriate assay design tool with tailored parameters, researchers can lay the foundation for designing SNP genotyping assays that meet their research goals effectively and accurately.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Raw Data Pre-processing

The raw sequencing data obtained from *Morus indica*, encompassing four distinct genotypes, underwent a meticulous quality control process. The initial quality assessment revealed. Subsequently, the raw data was subjected to quality filtering and adapter trimming to ensure data integrity and reliability.

4.2 Quality Checks and Reads

Quality assessment of the raw sequencing data retrieved from all four genotypes of *Morus indica* was conducted using FastQC and MultiQC toolkit. FastQC provided a comprehensive analysis, evaluating various aspects of the raw sequence data, including mean sequence quality per base position, per read, nucleotide content per base position, GC distribution, and other key metrics.

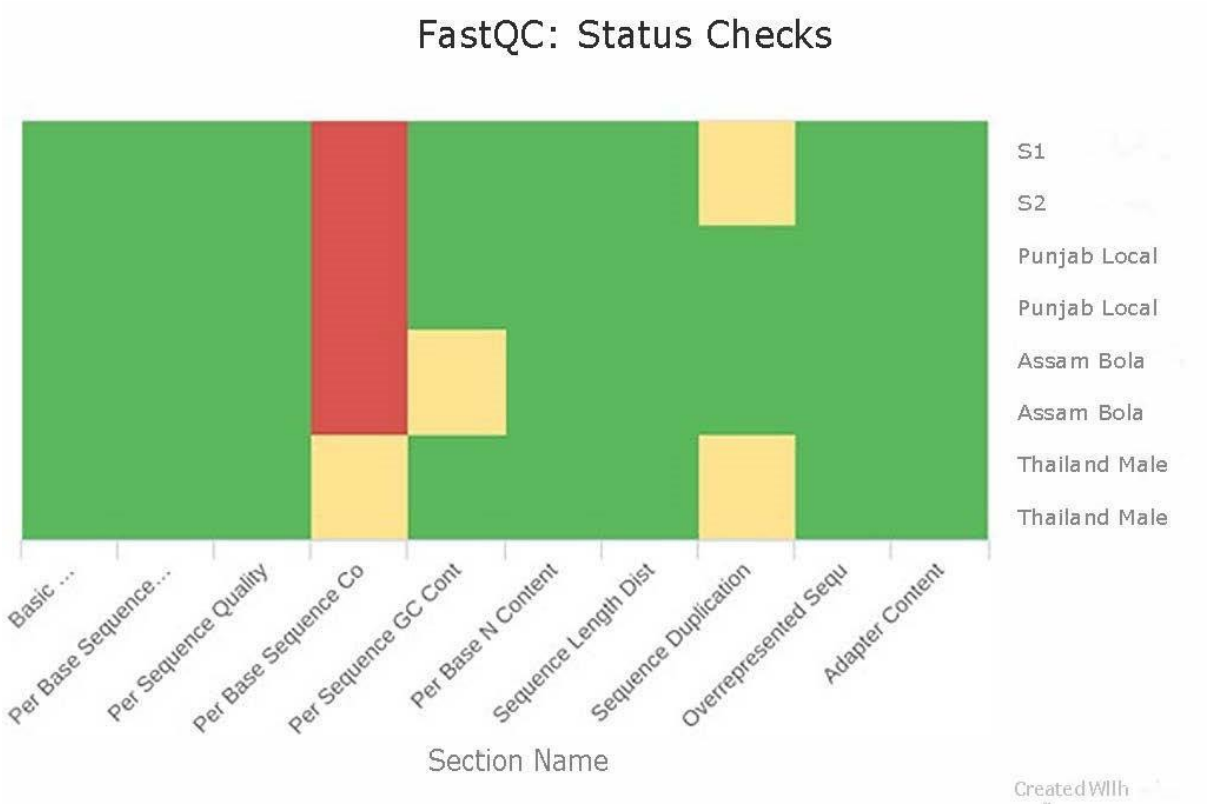


Figure 13: Quality Control Status for Sequencing Data

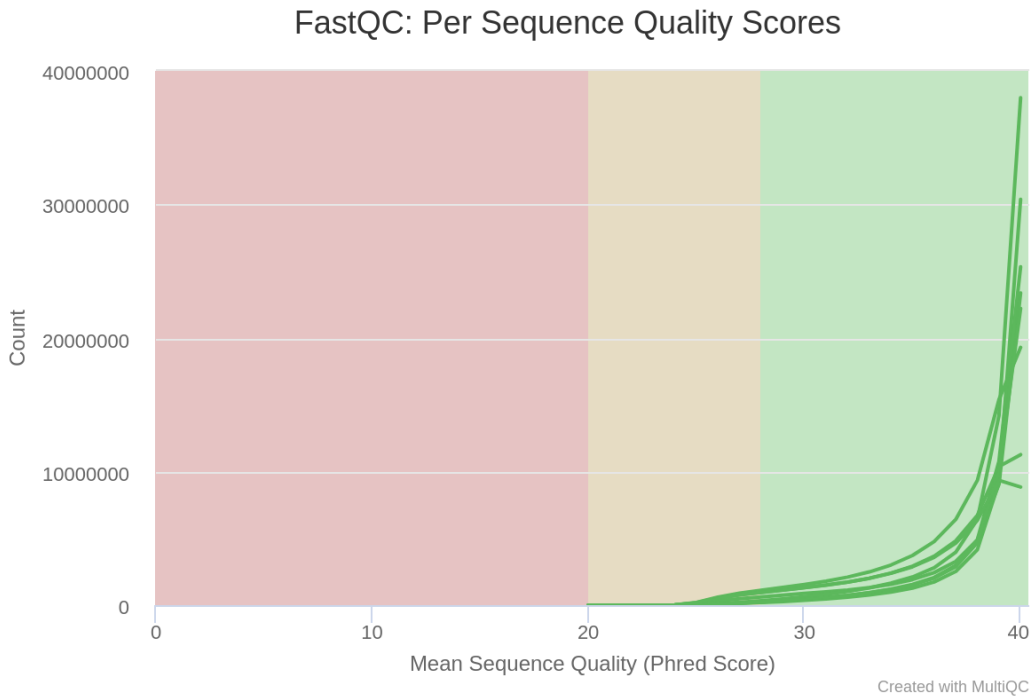


Figure 14: The image depicts the distribution of mean sequence quality scores, which are assigned to individual DNA sequences obtained from high-throughput sequencing devices. It illustrates that as the mean sequence quality (Phred Score) rises, the number of sequences sharply declines. This indicates a scarcity of sequences with higher quality scores. The graph effectively represents the predominance of good-quality scores among the sequences utilized in the study.

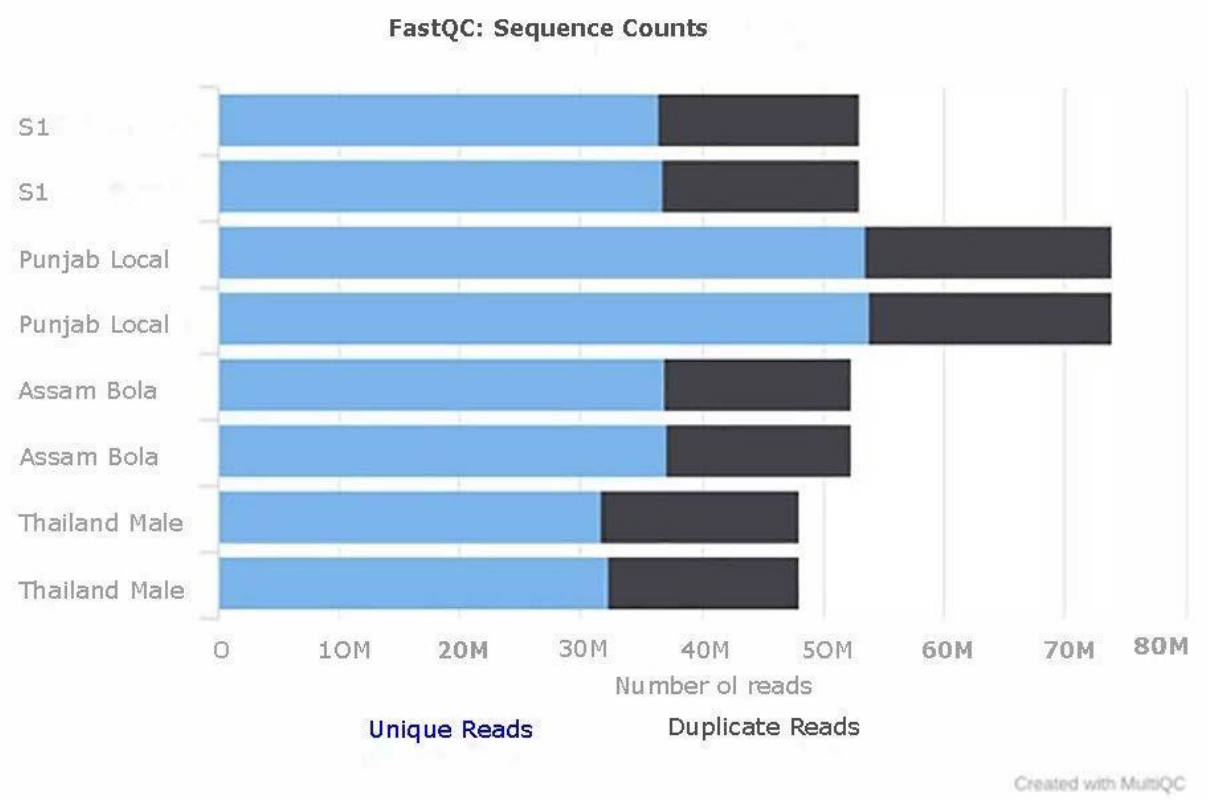


Figure 15: Sequence Counts of FastQC

The x-axis corresponds to sequence (genotypes), ranging from 0 to 80 million reads. The blue portion of each bar represents unique reads. These are individual sequences that occur only once in the dataset. Unique reads are essential for accurate data analysis, providing genuine information about the sample. The black portion of each bar represents duplicate reads. These are sequences that appear more than once in the dataset. Duplicate reads can arise due to technical artifacts during sequencing or PCR amplification biases. Assessing the balance between unique and duplicate reads is crucial for data quality. High-quality data should have a reasonable proportion of unique reads and minimal duplication. DNA sequencing generates millions of reads, resulting in gigabytes of data. Efficient analysis tools are necessary to make sense of this high-throughput sequencing (HTS) data.

4.3 Gapped Alignment

In the present investigation, we conducted a proficient alignment mapping of *Morus indica* exome reads, which were pre-processed, to the genome utilizing the Burrows-Wheeler Aligner (BWA). By employing BWA-backtrack, BWA-MEM, and BWA-SW algorithms, our methodology ensured accurate alignments within an edit distance of 2 to the query sequence. Furthermore, we assessed the data using Bowtie2, a highly efficient tool recognized for its proficiency in aligning short reads to extensive genomes, including that of *Morus indica*. The resulting SAM-formatted output is compatible with various downstream tools, especially the Genome Analysis Toolkit (GATK), facilitating subsequent comparative genomics studies. This strategic alignment lays the foundation for in-depth analyses, such as variant calling, offering valuable insights into the genomic intricacies of *Morus indica*. Importantly, all samples showcased robust alignment, each achieving a notable alignment rate of 85% or higher.

Table 3: The table clearly shows the genotypes and their corresponding alignment percentages

Sample ID	Genotypes	Alignment Percentage (%)
SRR14506990	S1	87%
SRR14506999	Punjab Local	85%
SRR14507001	Assama Bola	86%
SRR14507002	Thailand Male	85%

4.4 Sam to Bam conversion

Improved the alignment output to ensure high-quality alignments and reduce the risk of false variant calls in the genomic data of *Morus indica* (mulberry). We utilized popular software such as SAM tools, Genome Analysis Toolkit (GATK), and Picard to perform various operations on the SAM/BAM files generated from the alignment of mulberry from twenty *Morus indica* datasets. These operations involved sorting, recalibrating quality scores, marking PCR and optical duplicates, realigning indels, and filtering reads. SAM tools were particularly important for converting the alignment output to BAM format, as well as for sorting and merging BAM files. The choice of the SAM/BAM format was based on its simplicity, flexibility, and compatibility with data from different sequencing platforms. The binary BAM format, known for its small size and positional indexing, enabled quick retrieval of

alignments in specific regions, ensuring efficient downstream processing. This strategic use of SAM tools demonstrates an integrated approach to the thorough examination and analysis of large-scale genomic datasets, laying a strong groundwork for subsequent analyses, including variant calling and comprehensive exploration of the genomic landscape of *Morus indica*.

4.5 Variant Processing

The study successfully processed the variants using a combination of PICARD and GATK. PCR duplicates were identified using PICARD, while local re-alignment and base quality recalibration were performed using GATK with appropriate parameters. Sort Sam and building BAM indexes were also executed using PICARD. Once the covariates were analyzed, the variants were called.

4.6 Variant Calling

Following the variant processing steps, the identified variants were called utilizing SAMtools for pileup of mapped data and base counting, followed by variant calling using GATK. The total variants were identified, with a specific focus on SNPs. Subsequently, the SNPs and indels were extracted and subjected to filtering. The filtered SNPs and indels underwent annotation using snpEFF, a tool adept at annotating variants and predicting their effects on known genes. The analysis revealed comprehensive insights into the identified SNPs, facilitated by snpEFF's efficient annotation capabilities. Leveraging its compatibility with various genomes and rapid processing speed, snpEFF proved to be a valuable asset in annotating many mutations per minute.

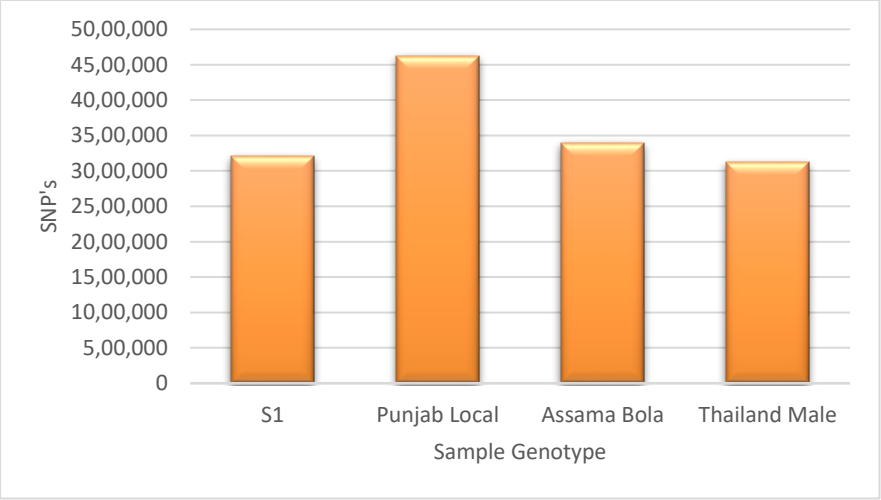


Figure 16: The graph represents the sample types and the corresponding SNP generated concerning the sample

However, managing the sheer complexity of the data can pose significant challenges, especially when dealing with millions of SNPs generated across multiple individuals. To simplify this intricate dataset, it often turns to generating consensus sequences. These consensus sequences distill the wealth of SNP information into a more manageable form by representing the most common nucleotide observed at each position across all individuals. Typically visualized with the x-axis denoting different sample genotypes and the y-axis indicating SNP positions, consensus sequences provide a simplified overview of genetic variation, with lines or curves illustrating the predominant nucleotide at each position. By focusing on these consensus sequences, we could more effectively analyze the dataset, identifying prevalent variations, uncovering patterns, and elucidating associations with traits or diseases. Ultimately, this approach facilitates a deeper understanding of genetic diversity within populations, enabling researchers to extract meaningful insights from complex genomic data.

4.7 Identifying SSRs

Results obtained from the SSR identification process concerning mulberry genetics would provide insights into the distribution and characteristics of microsatellite markers within the mulberry genome, which could be valuable for various genetic studies, breeding programs, and conservation efforts related to mulberry species.

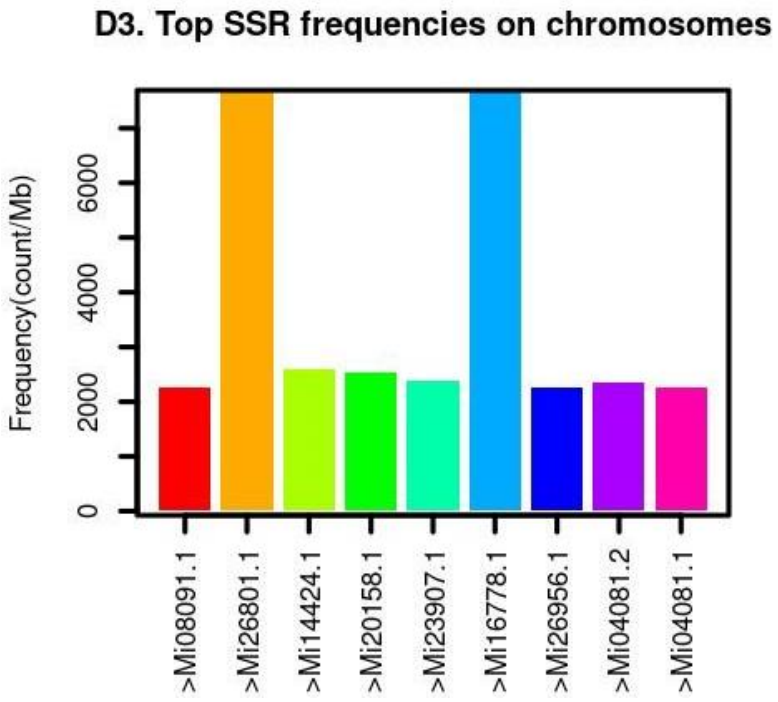


Figure 17: The graph displays the frequency count of SSRs (Simple Sequence Repeats) on different chromosome segments or types. The Y-axis represents the frequency in counts per megabase (Count/Mb), ranging from 0 to 6000. The X-axis labels correspond to various chromosome segments, such as "JN108091.1," "JN126801.1," "JN144241," and so forth. There are eight bars in total, each of a different color. The bar labeled "JN167781" stands out with the highest frequency count, reaching up to 6000.

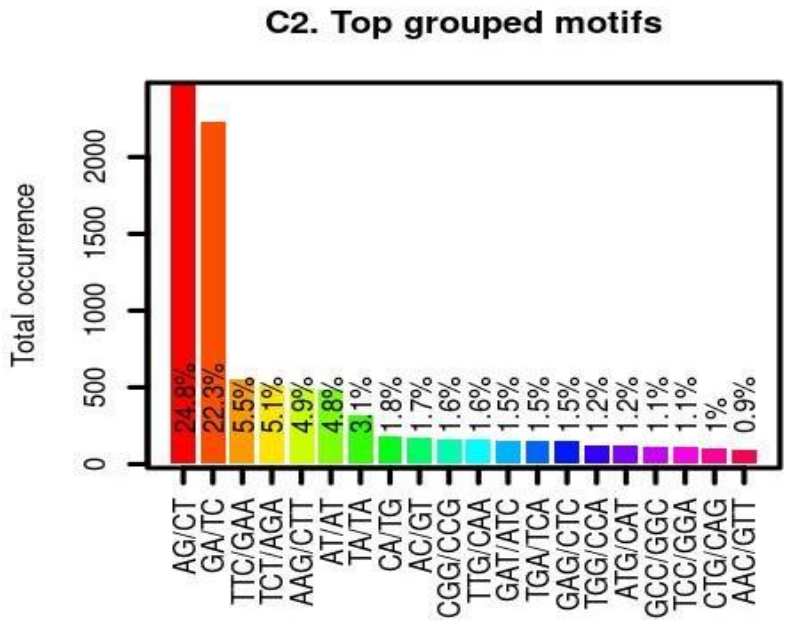


Figure 18: The Picture depicts a bar graph showing the top-grouped motifs in DNA sequences. These motifs are short, recurring patterns in DNA that are presumed to have a biological function. The graph highlights the total occurrence and diversity of these motifs.

4.8 Database development

The Indian Mulberry Linkage Mapping Database (IMLM-DB) represents a significant milestone in the genetic and breeding research of *Morus indica*, a species with diverse therapeutic and economic importance. The database focuses on Single Nucleotide Polymorphisms (SNPs) and Simple Sequence Repeats (SSRs), crucial molecular markers increasingly utilized in genetic diversity analysis, marker-assisted selection, genome and QTL mapping, and conservation genetics. A genotyping array comprising of 25K SNPs has been developed to identify markers associated with specific traits crucial for mulberry leaf production enhancement. This array has been designed through comprehensive bioinformatic analysis of the whole genome and transcriptome data obtained from the network project NW1. Approximately 350 germplasm accessions have been genotyped using the developed array. Promising trait donor genotypes have been identified through rigorous analysis of genotypic data, facilitating targeted breeding efforts aimed at enhancing yield potential, drought adaptation, and disease resistance in mulberry plants. A comprehensive database of SSR and SNP markers has been established, providing researchers with valuable genetic information for various applications in mulberry breeding and conservation. SNP markers identified through whole genome alignment of 20 contrasting lines have been cataloged, contributing to the development of robust DNA-based molecular

markers for accelerating trait introgression and improving mulberry leaf production. The project has provided genotyping support to other subprojects, enabling the demonstration of effective Linkage Disequilibrium (LD) decay and the construction of a dense linkage map using codominant markers like SSRs and SNPs.

https://mulbery.netlify.app/

4.9 Consensus Generation

The VCF files yielded an extensive dataset with 5 million SNPs. To distill the most pertinent SNP variations within a sequence, consensus generation was undertaken. Through this process, a contig file was produced, encapsulating the most specific SNP variations identified from the dataset.

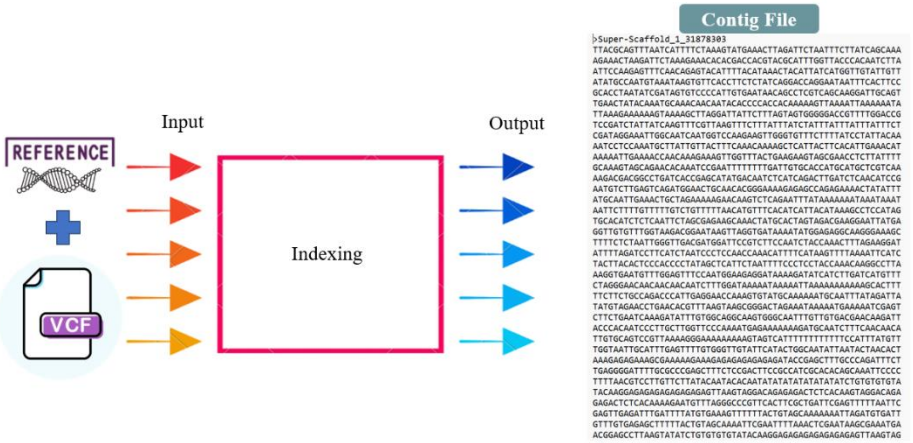


Figure 19: Contig file generated after consensus generation

4.10 Primer Design

The Primer3 Primer design suite was harnessed to craft primers for amplification. Leveraging the start and end regions of the contig file, primers were meticulously designed to target specific regions within the sequence.

Scaffold	Gene		Start	End	REF	ALT	Primers
Super-Scaffold_1_31878303	ID=Mi00003	gene	21478	31270	A	C	AGGGTTTAATGTTTGAAAGACGAG
Super-Scaffold_1_31878303	ID=Mi00003	gene	21478	31270	T	A	CTAAAGGCTCGAGGAGAGGAG
Super-Scaffold_1_31878303	ID=Mi00004	gene	31934	34104	A	T	CCAATCCATTTATTATGGACTCAA
Super-Scaffold_1_31878303	ID=Mi00005	gene	34262	37168	G	A	CCGACAATAAGTTCTCCTCTCTCT
Super-Scaffold_1_31878303	ID=Mi00006	gene	38384	40920	C	G	GACTTGGATTTTCGAGCGAC
Super-Scaffold_1_31878303	ID=Mi00007	gene	41411	47108	G	T	AATCGGAGAAGAACAAGAAGAAGA
Super-Scaffold_1_31878303	ID=Mi00008	gene	47732	51530	A	G	AAGTTCCAACTCGCAGGAAA
Super-Scaffold_1_31878303	ID=Mi00008	gene	47732	51530	G	A	CAAGAGGGACAGAGTACAGACCTT
Super-Scaffold_1_31878303	ID=Mi00008	gene	47732	51530	T	C	ACGCTCAGCCTCATGGTAAC
Super-Scaffold_1_31878303	ID=Mi00008	gene	47732	51530	C	A	TAATGAGCGGCTACAAAATTAGG
Super-Scaffold_1_31878303	ID=Mi00008	gene	47732	51530	A	G	TAATGAGCGGCTACAAAATTAGG
Super-Scaffold_1_31878303	ID=Mi00009	gene	51111	56248	G	T	GGTTGATGACATAATCAATCAAGAA
Super-Scaffold_1_31878303	ID=Mi00012	gene	67310	74445	T	C	TGAACTGTTAATTACTATGCTCTGCTC
Super-Scaffold_1_31878303	ID=Mi00012	gene	67310	74445	T	A	CCTGTATTGTACTACTCTTCTTCTTTGC
Super-Scaffold_1_31878303	ID=Mi00012	gene	67310	74445	A	T	CACGACTAAAAAAAAAAAAGAAACTCAG
Super-Scaffold_1_31878303	ID=Mi00012	gene	67310	74445	C	T	CTGCAGTTTATATTGTCTCTTTGATATTTT
Super-Scaffold_1_31878303	ID=Mi00012	gene	67310	74445	G	C	CTCTACTTTTAACTTACATGATTCTTTGGA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	A	G	GGAATCAGACGATGCTGACA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	C	T	GGAATCAGACGATGCTGACA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	G	A	AGTCTCAGATGTAGAAACTTAGGGTGA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	T	C	AGTCTCAGATGTAGAAACTTAGGGTGA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	T	G	AGTCTCAGATGTAGAAACTTAGGGTGA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	A	G	AGTCTCAGATGTAGAAACTTAGGGTGA
Super-Scaffold_2_30410261	ID=Mi02010	gene	7336	14889	A	G	AGTCCTGAGCTGTAGAGCTGTTTT

Figure 20: The picture depicts the primers that are designed using the primer3 design suite

https://docs.google.com/spreadsheets/d/1QoLKp1zCKGdwNFQH9RIrk3f0bnHzlDu2/edit?usp=sharing&ouid=114613627565011743972&rtpof=true&sd=true

4.11 Genotype/phenotype relation

Gene Mi00022 (GO:0046872 - Response to Water Deprivation), Alleles: Reference allele T, Alternative allele A, Breed: Punjab Local, Phenotype Associated: Drought. The presence of the alternative allele A in gene Mi00022 in the Punjab Local breed suggests a potential association with drought tolerance or response to water deprivation. This genotype-phenotype correlation implies that individuals carrying the alternative allele may exhibit enhanced resilience or adaptation to drought conditions compared to those with the reference allele T. Gene Mi00043 (GO:0009399 - Nitrogen Fixation), Alleles: Reference allele A, Alternative allele T, Breed: Assama Bola, Phenotype Associated: NUE (Nutrient Use Efficiency). The alternative allele T in gene Mi00043, found in the Assama Bola breed, is linked to Nitrogen Fixation, which is associated with Nutrient Use Efficiency (NUE). This genotype-phenotype correlation suggests that individuals carrying the alternative allele may possess traits conducive to efficient nitrogen utilization, potentially leading to improved NUE in agricultural settings. Gene Mi00041 (GO:0016168 - Chlorophyll Binding), Alleles: Reference allele A, Alternative allele G, Breed: S1, Phenotype Associated: Leaf yield. The alternative allele G in gene Mi00041, observed in the S1 breed, is associated with Chlorophyll Binding, a process essential for photosynthesis and leaf development. The correlation with Leaf yield suggests that individuals carrying this allele may exhibit traits conducive to higher leaf productivity, potentially leading to increased overall yield in agricultural contexts. These genotype-phenotype correlations highlight the intricate relationships

between genetic variations and observable traits in plant populations. Understanding such correlations is essential for targeted breeding efforts aimed at enhancing desirable traits, such as drought tolerance, nutrient use efficiency, and yield, to improve agricultural sustainability and productivity.

Scaffold	Start	End	Gene	Ontology	Pathway related	Ref	Alt	Breed	Phenotype associated
Super-Scaffold_1_31878303	144943	148886	Mi00022	GO:0046872	Response to Water Deprivation	T	A	Punjab Local	Drought
Super-Scaffold_1_31878303	312334	322231	Mi00043	GO:0009399	Nitrogen Fixation	A	T	Assama Bola	NUE
Super-Scaffold_1_31878303	285619	291110	Mi00041	GO:0016168	Chlorophyll Binding	A	G	S1	Leaf yield
Super-Scaffold_1_31878307	45143	45236	ID=Mi00007	GO:0005634	Necrosis	A	G	Punjab Local	Root rot
Super-Scaffold_1_31878303	67310	74445	ID=Mi00012	GO:0015977	Carbon Fixation	G	A	Thailand male	Leaf yield

Figure 21: Gene Information and Phenotype Associations

https://docs.google.com/spreadsheets/d/1uGmqXDQL1pjPTfvxf_xt9GQMKrV3WfnxDhvCjfyePwQ/edit#gid=0

When plants face water scarcity, known as drought stress, it triggers complex changes in their physiology. This includes alterations in leaf water levels and adjustments to cope with the lack of water. Drought stress reduces photosynthesis and worsens with severe stress, affecting plant growth and yield. Certain plants tolerate drought better, showing fewer negative effects on their membranes and maintaining cell stability. These physiological adjustments help plants endure drought conditions, but severe stress can still reduce crop yield. Breeding for drought tolerance is crucial for mitigating these effects and ensuring food security amidst changing climate conditions.

In farming, nitrogen fixation by plants like legumes turns atmospheric nitrogen into usable forms, enriching soil fertility and reducing the need for synthetic fertilizers. Nitrogen use efficiency (NUE) gauges how well crops utilize nitrogen. This balance influences agricultural sustainability, impacting environmental health. Nitrogen fixation can enhance NUE by supplementing nitrogen supply, lessening reliance on synthetic fertilizers, and curbing environmental harm. Yet, the interaction between nitrogen fixation and NUE hinges on various factors like crop type and management practices. Understanding this link guides sustainable nitrogen management, vital for balanced agricultural productivity and environmental stewardship.

The connection between chlorophyll binding and leaf yield is crucial for a plant's ability to thrive through photosynthesis. Chlorophyll, essential for this process, captures light energy for plant growth. Leaf characteristics like size and chlorophyll content affect how efficiently plants can absorb and use light. Genetic differences in these traits impact yield and seed quality. For example, leaf shape and size influence how well plants can distribute light and utilize it effectively. These traits are controlled by

complex genetic factors, as seen in studies identifying specific genetic regions linked to leaf traits and chlorophyll content.

Necrosis, the death of plant tissue, often caused by diseases like root rot, severely impacts plant health and productivity. Root rot, typically triggered by pathogens like Phytophthora, attacks plant roots, hindering water and nutrient absorption, and harming growth and yield. Moreover, it worsens soil erosion and nutrient loss. On the other hand, carbon fixation, where plants convert carbon dioxide into usable forms, aids growth and soil fertility. It indirectly boosts leaf yield by enhancing nutrient availability and reducing fertilizer needs. However, the link between carbon fixation and yield is complex, varying with plant type, environment, and farming methods. Understanding these connections is vital for sustainable agriculture, ensuring healthier plants and higher yields.

4.12 Phenotype/phenotype relation

In this phenotype-phenotype correlation matrix, we observe intriguing relationships among the factors of Drought, Leaf Yield, NUE (Nutrient Use Efficiency), and Root Rot, shedding light on their interconnectedness and potential implications for plant physiology and agriculture. The strong negative correlation between Drought and NUE (-0.95) underscores a significant relationship. As drought conditions escalate, there is a notable decline in nutrient use efficiency. This suggests that genetic traits involved in drought response may influence how efficiently plants utilize nutrients, highlighting the intricate interplay between environmental stressors and physiological processes. The robust negative correlation between Leaf Yield and Root Rot (-1) reveals a compelling association. As leaf yield increases, there's a corresponding decrease in root rot incidence, and vice versa. This implies that genetic factors influencing leaf productivity might also confer resistance or susceptibility to root pathogens, or conversely, root health might affect the plant's capacity to produce leaves. This intricate relationship underscores the importance of considering both above-ground and below-ground traits in crop improvement efforts. While weaker, the correlations between other factors hint at additional phenotype-phenotype relationships. Although less pronounced, these associations may still offer valuable insights into how different traits interact and influence each other under varying environmental conditions or genetic backgrounds.

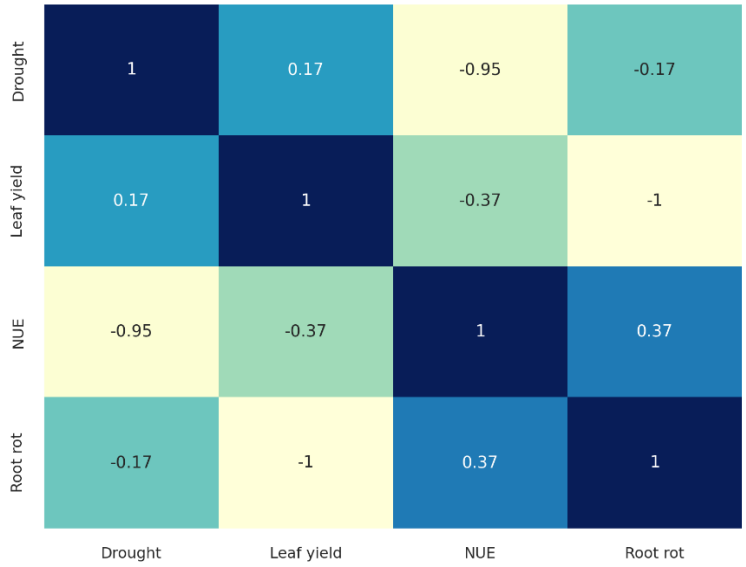


Figure 22: Correlation Matrix of Agricultural Factors

4.13 Genotyping Chip Generation

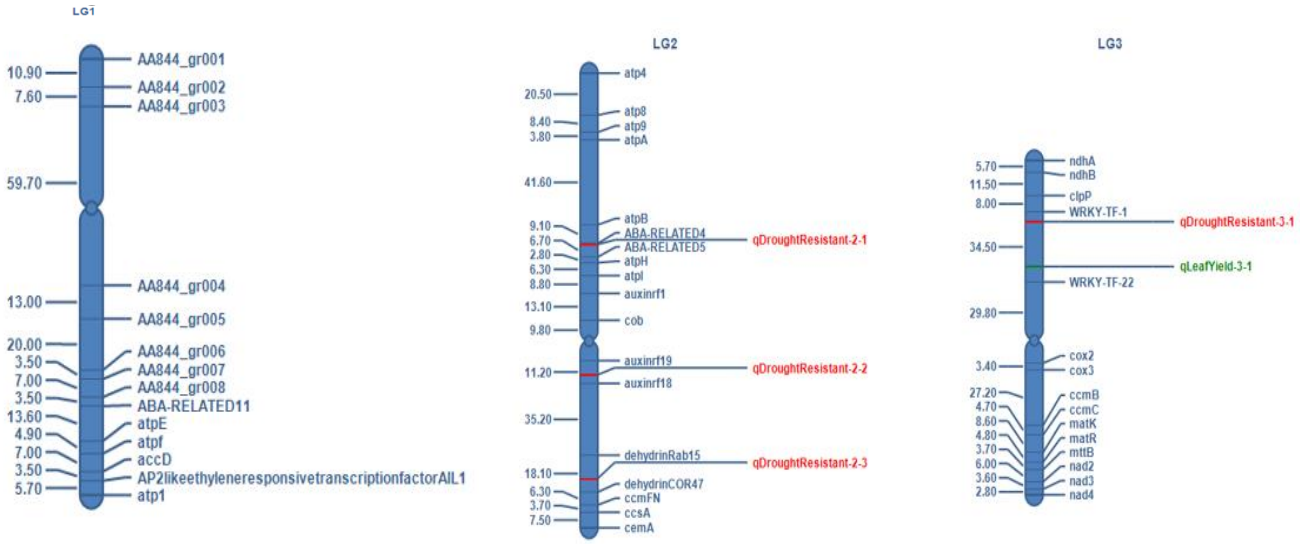
The table below contains data related to genotyping, particularly focusing on single nucleotide polymorphisms (SNPs) generated using the Illumina platform. The table includes various columns such as Illumina ID, Name, Quality, Illumina Strand, SNP, Address A ID, Primers, #genomeBuildk, Ploidy, Species, and Source. Quality Column Indicates the quality of SNPs, categorized into "High" and "Very High," suggesting the reliability and accuracy of the SNP data. The SNP Column Lists different types of single nucleotide polymorphisms, which are the variations at a single position in a DNA sequence among individuals. Ploidy Column All entries are marked as "polyploid," indicating that the organisms being studied have multiple sets of chromosomes, typical in many plant species. Species Column specifies "*Morus indica*6K_ARRAY," indicating that the genotyping chip used is designed specifically for the species *Morus indica* or a related organism, and it's tailored to interrogate approximately 6000 SNPs. The SNPs generated from this data are likely specific to *Morus indica* or a closely related species, and they have been obtained using a genotyping chip designed for this species. The "Source" column might provide additional information on the origin or sample type of the genetic material used for genotyping.

Assay Format	Infinium HD Ultra												
Date Manufactured	2/10/201												
Loci Count	7098												
[Assay]													
IlmnID	Name	Quality	IlmnStrand	SNP	AddressA_ID	Primers	GenomeBuil	Ploidy	Species	Source	rceVer	rceStr	Source seq
10001132-1_T_R_2584582511	10001132	High	TOP	T/G	32715340	ACACAAATCCGAATTTCTTTTGAT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10003037-1_B_F_2584581102	10003037	High	BOT	A/T	64731428	GCAACCCATCATCAGTACAAGTAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10012489-1_B_F_2584582512	10012489	High	BOT	C/T	16652430	GCAACCCATCATCAGTACAAGTAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10015956-1_T_F_2584581103	10015956	High	TOP	T/C	30807380	TATTAATTATTACCACGCACGACTACC	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	CCGAAGACGGTCACCGAAGACGGTCAAAT
10017499-1_T_R_2584583488	10017499	High	TOP	C/T	62801301	GAAGCCAGTCTTCTTGATCTTGTT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10017869-1_T_R_2584578901	10017869	High	TOP	G/C	20676421	AGGGTTTAATGTTTGAAAGACGAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10022789-1_B_R_2584583485	10022789	High	BOT	T/C	46806452	GTCTTCCAAAACGAGGTTAGCTT	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	CCGAAGACGGTCACCGAAGACGGTCAAAT
10022933-1_T_R_2584578902	10022933	High	TOP	G/A	12697319	CGTTGATTCACTTACGAATCTGAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10032781-1_T_F_2584578903	10032781	High	TOP	A/T	18717461	TTTTTTTTTTTCCGCTACGAAAG	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	CCGAAGACGGTCACCGAAGACGGTCAAAT
10060149-1_T_R_2584578904	10060149	High	TOP	G/T	15616470	TCCTCTCTGTATATTTCGAATGCTT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
10063204-1_T_R_2584583495	10063204	Very High	TOP	A/G	58633410	GAGCATAGATATTTTCAGCACCCT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
1006459-1_B_R_2584580858	1006459	High	BOT	T/C	51735320	ATACCATGACATGCATCCTATCAA	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	CCGAAGACGGTCACCGAAGACGGTCAAAT
10066953-1_T_R_2584580859	10066953	High	TOP	A/C	12633467	CTTCTTTTGCAAATGGTGGTAAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	CCGAAGACGGTCACCGAAGACGGTCAAAT
1007340-1_B_R_2584580861	1007340	High	BOT	G/A	15678462	CTTCTTTTGCAAATGGTGGTAAG	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	CCGAAGACGGTCACCGAAGACGGTCAAAT
10076803-1_B_R_2584583491	10076803	High	BOT	A/T	66764497	ATCATGATTCAAATTAGAGAACGAA	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10077150-1_T_R_2584578876	10077150	High	TOP	C/T	64806387	AGAACGAATCTAAATACCAAACGG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
1007944-1_T_R_2584584054	1007944	High	TOP	A/G	31777358	AGAACGAATCTAAATACCAAACGG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10080584-1_T_R_2584584056	10080584	Fail	TOP	G/C	18739434	CTACTGGTTACTGCTACTCCCTGTACT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10086999-1_B_F_2584578877	10086999	High	BOT	A/G	67754433	CTACTGGTTACTGCTACTCCCTGTACT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10087881-1_B_F_2584584057	10087881	High	BOT	T/A	30710309	ACACACACCTGATGAGGAAATAAT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10093765-1_T_F_2584583137	10093765	Fail	TOP	A/G	72716335	GCGTAGTCTTTTATTTCTGCGATT	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10094721-1_T_F_2584578878	10094721	Fail	TOP	T/A	32636508	GCTAGGTCTAGGTATGTTAGTTGAGTA TGA	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10098336-1_T_R_2584578879	10098336	High	TOP	T/A	56612374	GCAACCCATCATCAGTACAAGTAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10099158-1_T_R_2584583135	10099158	Very High	TOP	T/C	27646388	GCAACCCATCATCAGTACAAGTAG	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
1010821-1_B_R_2584583141	1010821	High	BOT	T/A	46717318	TATTAATTATTACCACGCACGACTACC	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
1010829-1_T_R_2584583142	1010829	High	TOP	T/C	48692415	GAAGCCAGTCTTCTTGATCTTGTT	7	polyploid	*Morus indica*	6K_ARRAY	1	BOT	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT
10108417-1_B_R_2584583145	10108417	High	BOT	A/C	27742418	AGGGTTTAATGTTTGAAAGACGAG	7	polyploid	*Morus indica*	6K_ARRAY	1	TOP	ATTATTATTATTATTATTATTGTTTTTTTAGGTTT

https://docs.google.com/spreadsheets/d/1DscSqLq64GhHlxUNnJu7mjt6i5EtUjA9/edit?usp=drive_link&ouid=104528999025534933339&rtpof=true&sd=true

4.14 Linkage Mapping

Drought Resistance, this trait is highly concentrated in LG-2, LG-4, and LG-5. Genes associated with drought resistance are likely clustered within these linkage groups. By identifying specific SNPs (single nucleotide polymorphisms) within these regions, researchers can pinpoint candidate genes responsible for drought tolerance. The gene IDs were homologically mapped with the gene names.



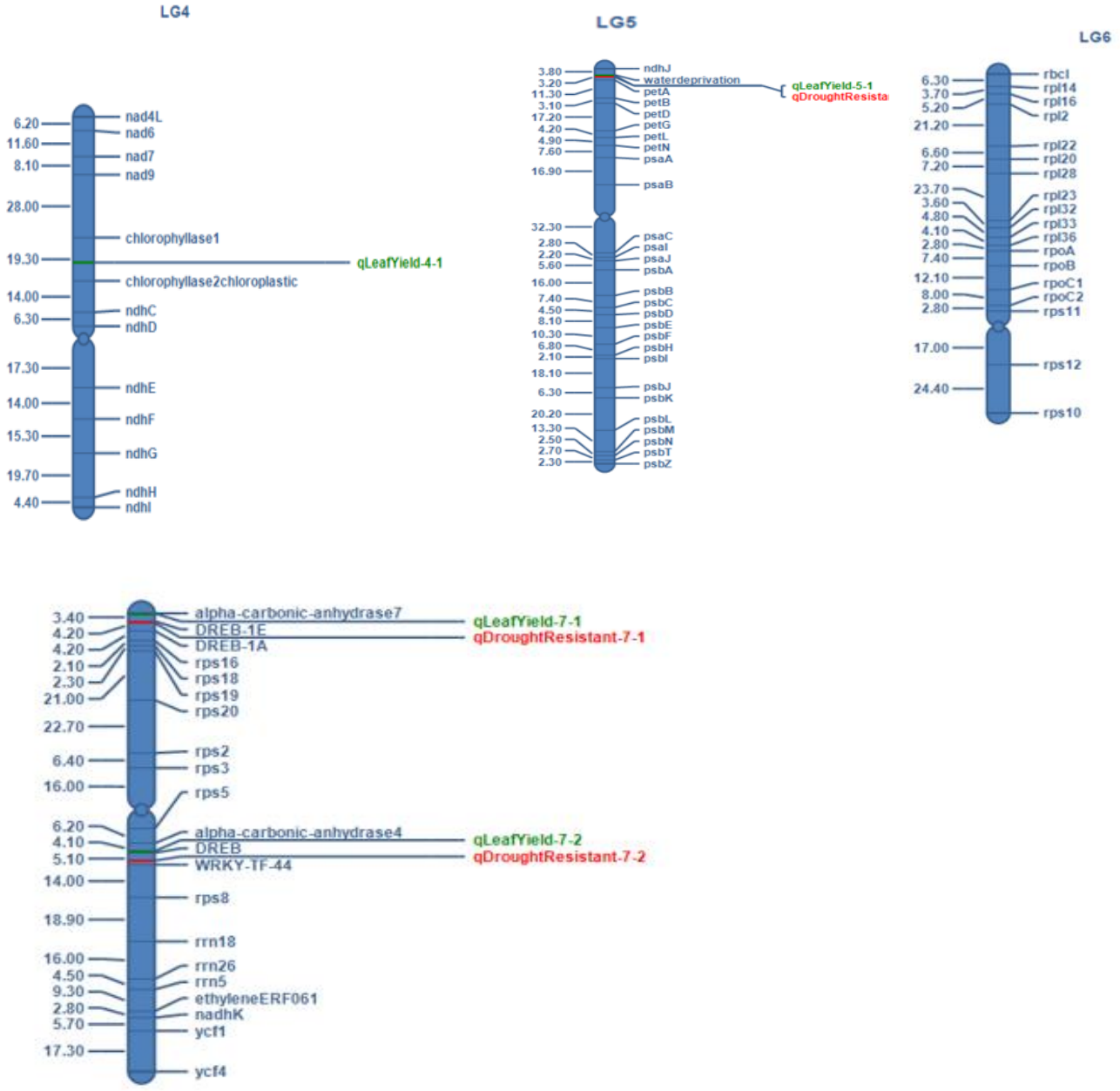


Figure 23: Marker-Trait association or linkage mapping

4.15 Utilizing Genotype-Phenotype Correlation and Specific SNPs for Informed Breeding Techniques in Agriculture

Farmers can use genotype-phenotype correlations to prioritize breeding efforts aimed at improving traits with a strong positive correlation, such as root rot resistance and nitrogen use efficiency. By selecting plants with favorable genotypes associated with these traits, breeders can develop cultivars better suited to combat root rot and optimize nitrogen utilization, leading to improved crop health and productivity. Given the positive correlation between nitrogen use efficiency and root rot resistance,

farmers can focus on breeding for cultivars that exhibit enhanced nitrogen use efficiency. This can involve selecting plants with specific SNPs associated with efficient nitrogen uptake, utilization, and assimilation, ultimately leading to improved crop performance and reduced environmental impact. Farmers can prioritize breeding goals based on the strength of genotype-phenotype correlations. For instance, while there may be a strong correlation between root rot resistance and nitrogen use efficiency, the low correlation between nitrogen use efficiency and leaf yield suggests that breeding efforts may need to be diversified to address multiple traits simultaneously, rather than focusing solely on increasing leaf yield. The genotype and phenotype mapping results, which include ontologies related to water deprivation, nitrogen fixation, necrosis, and carbon fixation, provide valuable insights into the underlying biological processes influencing trait variation. Farmers can use this information to inform breeding strategies aimed at enhancing resilience to water stress, improving nitrogen fixation efficiency, mitigating necrosis-related diseases, and optimizing carbon assimilation for improved yield and quality [154,155].

CHAPTER 5
CONCLUSION AND FUTURE ASPECT

5.1 Conclusion

In the culmination of this study, a comprehensive exploration of the genomic landscape of *Morus indica* has been undertaken, elucidating its multifaceted implications for agriculture and sericulture. Through meticulous data preprocessing, stringent quality assessment, comprehensive variant processing, and rigorous genotype-phenotype correlation analysis, invaluable insights into the genetic underpinnings of crucial traits essential for the successful cultivation of mulberry have been unearthed. This study has provided a comprehensive understanding of the genetic architecture governing traits such as drought resistance, nitrogen use efficiency, and leaf yield, paving the way for the development of targeted breeding techniques aimed at enhancing the resilience and productivity of mulberry cultivars. By establishing robust databases, conducting meticulous linkage mapping efforts, and pinpointing specific single nucleotide polymorphisms (SNPs), stakeholders have been equipped with indispensable resources to inform breeding strategies and expedite cultivar improvement programs.

Furthermore, this study underscores the critical importance of addressing existing research gaps, particularly in developing Indian-specific genomic resources and molecular markers. By bridging these lacunae, this study aims to empower farmers with actionable insights and tailored solutions for sustainable mulberry cultivation practices, fostering agricultural resilience and bolstering food security. Integrating genotype-phenotype and phenotype-phenotype relations serves as a linchpin in this endeavor, enabling farmers to make informed decisions and select breeds that exhibit optimal traits suited to their specific agro ecological contexts. Through this strategic approach, this study endeavors to optimize mulberry cultivation, maximize productivity, and elevate the economic prosperity of farming communities. This study serves as a cornerstone for continued advancements in mulberry genomics and its transformative application in agricultural innovation. The development of primers for chip development represents a pivotal step forward, poised to amplify the utilization of genomic information for precision breeding and cultivar enhancement. By leveraging the insights gleaned from this study and embracing emerging genomic paradigms, we are poised to catalyze a paradigm shift in mulberry cultivation, ensuring its enduring prominence and prosperity in the global agricultural landscape.

5.2 Future Aspects

The future holds exciting prospects for refining and advancing conservation and breeding strategies, all underpinned by the integration of comprehensive genetic insights. A key direction lies in the exploration of quantitative trait loci (QTL) analysis, a method aimed at identifying specific genomic regions associated with various agronomic traits crucial for mulberry cultivation. By deciphering the genetic basis of traits such as disease resistance, yield potential, and quality characteristics, breeders can target these regions for precise trait enhancement, thereby accelerating the development of superior mulberry cultivars. Moreover, the application of a metabolic approach presents a promising avenue for understanding the intricate biochemical pathways underlying trait variation in mulberry. By integrating metabolomic data with genomic information, researchers can unravel the complex interactions between genes and metabolites, providing deeper insights into the physiological processes governing mulberry growth, development, and productivity.

In parallel, the continued advancement of genetic monitoring programs holds immense potential for preserving the genetic diversity of mulberry populations. By incorporating real-time genomic data through high-throughput sequencing technologies and bioinformatics tools, conservation efforts can be bolstered, enabling timely interventions to mitigate threats and conserve the invaluable genetic resources harbored within mulberry germplasm. The expansion of marker-assisted breeding approaches, leveraging single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs), offers an efficient means to accelerate cultivar improvement. Marker-assisted selection (MAS) and genomic selection techniques empower breeders to harness genetic information for the identification and introgression of desirable traits into mulberry germplasm, thereby enhancing productivity, resilience, and quality characteristics. In the realm of functional genomics, ongoing studies are poised to elucidate the intricate biological mechanisms governing trait variation in mulberry. By integrating genomic data with transcriptomic, proteomic, and metabolomic datasets, researchers can unravel gene expression patterns, protein function, and metabolic pathways associated with critical agronomic traits. This comprehensive understanding paves the way for targeted manipulation of the mulberry genome to enhance desired traits, offering unprecedented opportunities for precision breeding. Moreover, the exploration of genome editing technologies, such as CRISPR-Cas9, heralds a new era of precise genetic

manipulation in mulberry cultivation. By enabling targeted gene knockout, gene editing for trait improvement, and allele replacement, genome editing holds immense potential for enhancing mulberry productivity, resilience, and quality characteristics in a precise and controlled manner. The integration of these diverse approaches and technologies promises to revolutionize mulberry cultivation, ensuring its sustainability and resilience in the face of evolving agricultural challenges. By embracing interdisciplinary collaboration and leveraging cutting-edge genetic tools, the future of mulberry agriculture is poised for remarkable advancements, driving innovation and prosperity in the global sericulture industry.

CHAPTER 6

REFERENCES

1) Encyclopedia. (1998). JAMA, 279(17), 1409. https://doi.org/10.1001/jama.279.17.1409-jbk0506-6-1 00/00544 Recent condition and trend of explosibility monitoring techniques for coal mine in China. (2000). Fuel and Energy Abstracts, 41(1), 57. https://doi.org/10.1016/0140-6701(00)95325-6

2) Yongkang Group. (2009). In Geological Formation Names of China (1866–2000) (pp. 1345-1345): Springer Berlin Heidelberg.

3) West Java. (2016). In 2015 Annual Competitiveness Analysis and Development Strategies for Indonesian Provinces (pp. 223-234): WORLD SCIENTIFIC.

http://ljournal.ru/wp-content/uploads/2017/03/a-2017-023.pdf. (2017). http://dx.doi.org/10.18411/a-2017-023

4) Identification of Sex-specific Molecular Markers in Barbel and Nile Carp using SCoT and ISSR Markers. (2021). Assiut Journal of Agricultural Sciences, 52(1), 60-73. https://doi.org/10.21608/ajas.2021.171568

5) Morus indica. (2022). CABI Publishing. https://doi.org/10.1079/cabicompendium.69105335

6) Acharya, R., Bagchi, T., & Gangopadhyay, D. (2022). Mulberry as a Valuable Resource for Food and Pharmaceutical Industries: A Review. In Medicinal Plants: IntechOpen.

7) Akhunov, E., Nicolet, C., & Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 119(3), 507-517. https://doi.org/10.1007/s00122-009-1059-5

8) Allen, A., Islamovic, E., Kaur, J., Gold, S., Shah, D., & Smith, T. J. (2011). Transgenic maize plants expressing the Totivirus antifungal protein, KP4, are highly resistant to corn smut. Plant Biotechnology Journal, 9(8), 857-864. https://doi.org/10.1111/j.1467-7652.2011.00590.x

9) Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L. A., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., Griffiths, S., Bentley, A. R., Alda, M., Jack, P., . . . Edwards, K. J. (2017). Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (Triticum aestivum). Plant biotechnology journal, 15(3), 390-401. https://doi.org/10.1111/pbi.12635

10) Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A., & Joets, J. (2004). BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics, 20(14), 2324-2326. https://doi.org/10.1093/bioinformatics/bth230

11) Asama, J., Suzuki, H., & Oiwa, T. (2016). Development of a magnetically suspended double-layered cylinder with 1-DOF active positioning control. Mechanical Engineering Journal, 3(1), 15-00544-00515-00544. https://doi.org/10.1299/mej.15-00544
12) Aversano, R., Ercolano, M., Caruso, I., Fasano, C., Rosellini, D., & Carputo, D. (2014). Molecular Tools for Exploring Polyploid Genomes in Plants. In The Role of Bioinformatics in Agriculture (pp. 73-97): Apple Academic Press.
13) Awasthi, A. K., Nagaraja, G. M., Naik, G. V., Kanginakudru, S., Thangavelu, K., & Nagaraju, J. (2004). Genetic diversity and relationships in mulberry (genus Morus) as revealed by RAPD and ISSR marker assays. BMC genetics, 5, 1-1. https://doi.org/10.1186/1471-2156-5-1
14) Baciu, E.-D., Baci, G.-M., Moise, A. R., & Dezmirean, D. S. (2023). A Status Review on the Importance of Mulberry (Morus spp.) and Prospects towards Its Cultivation in a Controlled Environment. Horticulturae, 9(4), 444. https://doi.org/10.3390/horticulturae9040444
15) Banerjee, R., Chattopadhyay, S., Saha, A. K., & Kumar, S. N. (2014). Association of morphological and biochemical features of mulberry lines with resistance to bacterial leaf spot. Archives Of Phytopathology And Plant Protection, 47(16), 2020-2032. https://doi.org/10.1080/03235408.2013.868693
16) Boro, H., Das, S., & Middha, S. K. (2021). The therapeutic potential and the health benefits of Morus indica Linn.: a mini review. Advances in Traditional Medicine, 21(2), 241-252. https://doi.org/10.1007/s13596-020-00544-5
17) Boro, H., Das, S., & Middha, S. K. (2021). The therapeutic potential and the health benefits of Morus indica Linn.: a mini review. Advances in Traditional Medicine, 21(2), 241-252. https://doi.org/10.1007/s13596-020-00544-5
18) Bukhari, R., & Sharma, A. (2022). IMPACT OF COVID-19 ON THE INDIAN SERICULTURE INDUSTRY. UTTAR PRADESH JOURNAL OF ZOOLOGY, 9-19. https://doi.org/10.56557/upjoz/2022/v43i203191
19) Burkart-Waco, D., Henry, I. M., Ngo, K., Comai, L., & Tai, T. H. (2016). Determining Mutation Density Using Restriction Enzyme Sequence Comparative Analysis (RESCAN). In Biotechnologies for Plant Mutation Breeding (pp. 305-321): Springer International Publishing.
20) Bus, A., Hecht, J., Huettel, B., Reinhardt, R., & Stich, B. (2012). High-throughput polymorphism detection and genotyping in Brassica napus using next-generation RAD sequencing. BMC genomics, 13, 281-281. https://doi.org/10.1186/1471-2164-13-281
21) Cai, M., Mu, L., Wang, Z.-L., Liu, J.-Y., Liu, T.-L., Wanapat, M., & Huang, B.-Z. (2019). Assessment of mulberry leaf as a potential feed supplement for animal feeding in P.R. China. Asian-Australasian journal of animal sciences, 32(8), 1145-1152.

https://doi.org/10.5713/ajas.18.0671

22) Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., da Silva, M. L., Bockelman, H., . . . Akhunov, E. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences of the United States of America, 110(20), 8057-8062. https://doi.org/10.1073/pnas.1217133110

23) Checker, V. G., Saeed, B., & Khurana, P. (2012). Analysis of expressed sequence tags from mulberry (Morus indica) roots and implications for comparative transcriptomics and marker identification. Tree Genetics & Genomes, 8(6), 1437-1450. https://doi.org/10.1007/s11295-012-0531-6

24) Checker, V. G., Saeed, B., & Khurana, P. (2012). Analysis of expressed sequence tags from mulberry (Morus indica) roots and implications for comparative transcriptomics and marker identification. Tree Genetics & Genomes, 8(6), 1437-1450. https://doi.org/10.1007/s11295-012-0531-6

25) Cheema, J., Ellis, T. H. N., & Dicks, J. (2010). THREaD Mapper Studio: a novel, visual web server for the estimation of genetic linkage maps. Nucleic acids research, 38(Web Server issue), W188-W193. https://doi.org/10.1093/nar/gkq430

26) Chen, C., You, L.-J., Abbasi, A. M., Fu, X., & Liu, R. H. (2015). Optimization for ultrasound extraction of polysaccharides from mulberry fruits with antioxidant and hyperglycemic activity in vitro. Carbohydrate Polymers, 130, 122-132. https://doi.org/10.1016/j.carbpol.2015.05.003

27) Chen, C., You, L.-J., Abbasi, A. M., Fu, X., Liu, R. H., & Li, C. (2016). Characterization of polysaccharide fractions in mulberry fruit and assessment of their antioxidant and hypoglycemic activities in vitro. Food & Function, 7(1), 530-539. https://doi.org/10.1039/c5fo01114k

28) Chen, C., Zhang, B., Fu, X., & Liu, R. H. (2016). A novel polysaccharide isolated from mulberry fruits (Murus alba L.) and its selenide derivative: structural characterization and biological activities. Food & Function, 7(6), 2886-2897. https://doi.org/10.1039/c6fo00370b

29) Chen, S., Huang, Z., Dai, Y., Qin, S., Gao, Y., Zhang, L., Gao, Y., & Chen, J. (2013). The development of 7E chromosome-specific molecular markers for Thinopyrum elongatum based on SLAF-seq technology. PloS one, 8(6), e65122-e65122. https://doi.org/10.1371/journal.pone.0065122

30) Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. Trends in plant

science, 15(2), 57-71. https://doi.org/10.1016/j.tplants.2009.12.003

31) Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Le Paslier, M.-C., Magdelenat, G., Gonthier, C., Couloux, A., Budak, H., Breen, J., Pumphrey, M., Liu, S., Kong, X., Jia, J., Gut, M., Brunel, D., . . . Feuillet, C. (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. The Plant cell, 22(6), 1686-1701. https://doi.org/10.1105/tpc.110.074187

32) Cifuentes, M., Eber, F., Lucas, M.-O., Lode, M., Chèvre, A.-M., & Jenczewski, E. (2010). Repeated polyploidy drove different levels of crossover suppression between homoeologous chromosomes in Brassica napus allohaploids. The Plant cell, 22(7), 2265-2276. https://doi.org/10.1105/tpc.109.072991

33) Claros, M. G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P., & Fernández-Pozo, N. (2012). Why assembling plant genome sequences is so challenging. Biology, 1(2), 439-459. https://doi.org/10.3390/biology1020439

34) Comai, L. (2005). The advantages and disadvantages of being polyploid. Nature Reviews Genetics, 6(11), 836-846. https://doi.org/10.1038/nrg1711

35) Dhanyalakshmi, K. H., & Nataraja, K. N. (2018). Mulberry (Morus spp.) has the features to treat as a potential perennial model system. Plant signaling & behavior, 13(8), e1491267-e1491267. https://doi.org/10.1080/15592324.2018.1491267

36) Dhanyalakshmi, K. H., & Nataraja, K. N. (2018). Mulberry (Morus spp.) has the features to treat as a potential perennial model system. Plant signaling & behavior, 13(8), e1491267-e1491267. https://doi.org/10.1080/15592324.2018.1491267

37) Duhan, N., Kaur, S., & Kaundal, R. (2023). ranchSATdb: A Genome-Wide Simple Sequence Repeat (SSR) Markers Database of Livestock Species for Mutant Germplasm Characterization and Improving Farm Animal Health. Genes, 14(7), 1481. https://doi.org/10.3390/genes14071481

38) 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. PloS one, 6(5), e19379-e19379. https://doi.org/10.1371/journal.pone.0019379

39) Endelman, J. B., & Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. Bioinformatics, 30(11), 1623-1624. https://doi.org/10.1093/bioinformatics/btu091

40) Faiz, A. U. H., & Faiz, L. Z. (2021). Phytochemical Screening, Antimicrobial Activity of Gemmotherapeutically White Mulberry (Morus Alba) Leaves. Journal of Bioresource Management, 8(2), 14-19. https://doi.org/10.35691/jbm.1202.0176

41) Farvid, M. S., Ding, M., Pan, A., Sun, Q., Chiuve, S. E., Steffen, L. M., Willett, W. C., & Hu, F. B. (2014). Dietary linoleic acid and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. Circulation, 130(18), 1568-1578. https://doi.org/10.1161/CIRCULATIONAHA.114.010236
42) Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., & Osborn, T. C. (2007). Genomic changes in resynthesized Brassica napus and their effect on gene expression and phenotype. The Plant cell, 19(11), 3403-3417. https://doi.org/10.1105/tpc.107.054346
43) Geng, X., Jiang, C., Yang, J., Wang, L., Wu, X., & Wei, W. (2016). Rapid Identification of Candidate Genes for Seed Weight Using the SLAF-Seq Method in Brassica napus. PloS one, 11(1), e0147580-e0147580. https://doi.org/10.1371/journal.pone.0147580
44) Ghorpade, K., Netam, V., Desai, N., & Shewale, V. (2022). ASCERTAINING ROLE AND IMPORTANCE OF ARTIFICIAL DIET AND BOTANICALS MEDIATED SERIFEED FOR THE STRENGTHENING OF SERICULTURE INDUSTRY. UTTAR PRADESH JOURNAL OF ZOOLOGY, 46-54. https://doi.org/10.56557/upjoz/2022/v43i223230
45) Giampieri, F., Tulipani, S., Alvarez-Suarez, J. M., Quiles, J. L., Mezzetti, B., & Battino, M. (2012). The strawberry: Composition, nutritional quality, and impact on human health. Nutrition, 28(1), 9-19. https://doi.org/10.1016/j.nut.2011.08.009
46) Gnanesh, B. N., Arunakumar, G. S., Tejaswi, A., Supriya, M., Manojkumar, H. B., & Devi, S. S. (2022). Characterization and Pathogenicity of Lasiodiplodia theobromae Causing Black Root Rot and Identification of Novel Sources of Resistance in Mulberry Collections. The plant pathology journal, 38(4), 272-286. https://doi.org/10.5423/PPJ.OA.01.2022.0005
47) Gnanesh, B. N., Mondal, R., G S, A., H B, M., Singh, P., M R, B., P, S., Burji, S. M., T, M., & V, S. (2023). Genome size, genetic diversity, and phenotypic variability imply the effect of genetic variation instead of ploidy on trait plasticity in the cross-pollinated tree species of mulberry. PloS one, 18(8), e0289766-e0289766. https://doi.org/10.1371/journal.pone.0289766
48) Grześkowiak, J., Łochyńska, M., & Frankowski, J. (2022). Sericulture in Terms of Sustainable Development in Agriculture. Problemy Ekorozwoju, 17(2), 210-217. https://doi.org/10.35784/pe.2022.2.23
49) Hamilton, J. P., & Robin Buell, C. (2012). Advances in plant genome sequencing. The Plant Journal, 70(1), 177-190. https://doi.org/10.1111/j.1365-313x.2012.04894.x
50) He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., Lee, T.-H., Wang, X., Cai, Q., Li, D., Lu, M., Liao, S., Luo, G., He, R., Tan, X., Xu, Y., Li, T., Zhao, A., Jia, L., . . . Xiang, Z. (2013). Draft genome sequence of the mulberry tree Morus notabilis. Nature communications, 4, 2445-2445. https://doi.org/10.1038/ncomms3445

51) Huang, H.-P., Ou, T.-T., & Wang, C.-J. (2013). Mulberry (sang shèn zǐ) and its bioactive compounds, the chemoprevention effects and molecular mechanisms in vitro and in vivo. Journal of traditional and complementary medicine, 3(1), 7-15. https://doi.org/10.4103/2225-4110.106535

52) Illahi, I., Mittal, V., & Sharmila, K. K. (2022). Integrated Pest and Disease Management (IPDM) Approach in Mulberry Sericulture of Jammu & Kashmir. Research Biotica, 4(3), 161-165. https://doi.org/10.54083/resbio/4.3.2022/161-165

53) Islam, M. S., Ahmed, F., & Rahman, M. M. (2023). Intercropping pulses with mulberry on sericulture productivity and profitability. Bangladesh Journal of Agriculture, 47(2), 107-118. https://doi.org/10.3329/bjagri.v47i2.63330

54) Ito, T., & Arai, N. (1966). Nutrition of the silkworm, Bombyx mori—XI. Requirements for aspartic and glutamic acids. Journal of Insect Physiology, 12(7), 861-869. https://doi.org/10.1016/0022-1910(66)90036-9

55) Jain, M., Bansal, J., Rajkumar, M. S., Sharma, N., Khurana, J. P., & Khurana, P. (2022). Draft genome sequence of Indian mulberry (Morus indica) provides a resource for functional and translational genomics. Genomics, 114(3), 110346. https://doi.org/10.1016/j.ygeno.2022.110346

56) Jain, M., Bansal, J., Rajkumar, M. S., Sharma, N., Khurana, J. P., & Khurana, P. (2022). Draft genome sequence of Indian mulberry (Morus indica) provides a resource for functional and translational genomics. Genomics, 114(3), 110346. https://doi.org/10.1016/j.ygeno.2022.110346

57) Jain, M., Bansal, J., Rajkumar, M. S., Sharma, N., Khurana, J. P., & Khurana, P. (2022). Draft genome sequence of Indian mulberry (Morus indica) provides a resource for functional and translational genomics. Genomics, 114(3), 110346. https://doi.org/10.1016/j.ygeno.2022.110346

58) Jan, B., Parveen, R., Zahiruddin, S., Khan, M. U., Mohapatra, S., & Ahmad, S. (2021). Nutritional constituents of mulberry and their potential applications in food and pharmaceuticals: A review. Saudi journal of biological sciences, 28(7), 3909-3921. https://doi.org/10.1016/j.sjbs.2021.03.056

59) Jan, B., Parveen, R., Zahiruddin, S., Khan, M. U., Mohapatra, S., & Ahmad, S. (2021). Nutritional constituents of mulberry and their potential applications in food and pharmaceuticals: A review. Saudi journal of biological sciences, 28(7), 3909-3921. https://doi.org/10.1016/j.sjbs.2021.03.056

60) Jansen, J., de Jong, A. G., & van Ooijen, J. W. (2001). Constructing dense genetic linkage maps.

Theoretical and Applied Genetics, 102(6-7), 1113-1122. https://doi.org/10.1007/s001220000489

61) Jiao, W.-B., & Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. Current Opinion in Plant Biology, 36, 64-70. https://doi.org/10.1016/j.pbi.2017.02.002

62) Jones, S. Identification of spectral disease signatures and resistant QTL for charcoal rot infection in soybean Iowa State University]. http://dx.doi.org/10.31274/etd-180810-5913

63) Ju, W.-T., Kwon, O. C., Kim, H.-B., Sung, G.-B., Kim, H.-W., & Kim, Y.-S. (2018). Qualitative and quantitative analysis of flavonoids from 12 species of Korean mulberry leaves. Journal of food science and technology, 55(5), 1789-1796. https://doi.org/10.1007/s13197-018-3093-2

64) Kaur, S., Francki, M. G., & Forster, J. W. (2011). Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. Plant Biotechnology Journal, 10(2), 125-138. https://doi.org/10.1111/j.1467-7652.2011.00644.x

65) Kerstin, M. (2006). Food and Agriculture Organization of the United Nations (FAO). In Max Planck Encyclopedia of Public International Law. Oxford University Press. https://doi.org/10.1093/law:epil/9780199231690/e462

66) Kim, D., & Kang, K.-H. (2022). Anti-Inflammatory and Anti-Bacterial Potential of Mulberry Leaf Extract on Oral Microorganisms. International journal of environmental research and public health, 19(9), 4984. https://doi.org/10.3390/ijerph19094984

67) Kim, D., & Kang, K.-H. (2022). Anti-Inflammatory and Anti-Bacterial Potential of Mulberry Leaf Extract on Oral Microorganisms. International journal of environmental research and public health, 19(9), 4984. https://doi.org/10.3390/ijerph19094984

68) Kim, D., & Kang, K.-H. (2022). Anti-Inflammatory and Anti-Bacterial Potential of Mulberry Leaf Extract on Oral Microorganisms. International journal of environmental research and public health, 19(9), 4984. https://doi.org/10.3390/ijerph19094984

69) Kusano, G., Orihara, S., Tsukamoto, D., Shibano, M., Coskun, M., Guvenc, A., & Erdurak, C. S. (2002). Five New Nortropane Alkaloids and Six New Amino Acids from the Fruit of Morus alba LINNE Growing in Turkey. Chemical and Pharmaceutical Bulletin, 50(2), 185-192. https://doi.org/10.1248/cpb.50.185

70) Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., Xiang, Z., Song, W., Ying, K., Zhang, M., Jiao, Y., Ni, P., Zhang, J., Li, D., Guo, X., Ye, K., Jian, M., Wang, B., Zheng, H., . . . Wang, J. (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. Nature genetics, 42(11), 1027-1030. https://doi.org/10.1038/ng.684

71) Le, H. D., Fallon, E. M., Kalish, B. T., de Meijer, V. E., Meisel, J. A., Gura, K. M., Nose, V.,

Pan, A. H., Bistrian, B. R., & Puder, M. (2013). The effect of varying ratios of docosahexaenoic acid and arachidonic acid in the prevention and reversal of biochemical essential fatty acid deficiency in a murine model. Metabolism: clinical and experimental, 62(4), 499-508. https://doi.org/10.1016/j.metabol.2012.10.003

72) Lee, J. S., Synytsya, A., Kim, H. B., Choi, D. J., Lee, S., Lee, J., Kim, W. J., Jang, S., & Park, Y. I. (2013). Purification, characterization and immunomodulating activity of a pectic polysaccharide isolated from Korean mulberry fruit Oddi (Morus alba L.). International Immunopharmacology, 17(3), 858-866. https://doi.org/10.1016/j.intimp.2013.09.019

73) Leitch, I. J., & Bennett, M. D. (1997). Polyploidy in angiosperms. Trends in Plant Science, 2(12), 470-476. https://doi.org/10.1016/s1360-1385(97)01154-0

74) Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R. G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., . . . Yu, S. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nature Biotechnology, 33(5), 524-530. https://doi.org/10.1038/nbt.3208

75) Li, T., Qi, X., Zeng, Q., Xiang, Z., & He, N. (2014). MorusDB: a resource for mulberry genomics and genome biology. Database : the journal of biological databases and curation, 2014, bau054. https://doi.org/10.1093/database/bau054

76) Li, W., Huang, L., & Gill, B. S. (2008). Recurrent deletions of puroindoline genes at the grain hardness locus in four independent lineages of polyploid wheat. Plant physiology, 146(1), 200-212. https://doi.org/10.1104/pp.107.108852

77) Liu, B., Vega, J. M., & Feldman, M. (1998). Rapid genomic changes in newly synthesized amphiploids ofTriticumandAegilops. II. Changes in low-copy coding DNA sequences. Genome, 41(4), 535-542. https://doi.org/10.1139/g98-052

78) Liu, C.-J., & Lin, J.-Y. (2014). Protective effects of strawberry and mulberry fruit polysaccharides on inflammation and apoptosis in murine primary splenocytes. Journal of Food and Drug Analysis, 22(2), 210-219. https://doi.org/10.1016/j.jfda.2014.01.015

79) Liu, Y., Salsman, E., Wang, R., Galagedara, N., Zhang, Q., Fiedler, J. D., Liu, Z., Xu, S., Faris, J. D., & Li, X. (2020). Meta-QTL analysis of tan spot resistance in wheat. Theoretical and Applied Genetics, 133(8), 2363-2375. https://doi.org/10.1007/s00122-020-03604-1

80) Liu, Y., Salsman, E., Wang, R., Galagedara, N., Zhang, Q., Fiedler, J. D., Liu, Z., Xu, S., Faris, J. D., & Li, X. (2020). Meta-QTL analysis of tan spot resistance in wheat. Theoretical and Applied Genetics, 133(8), 2363-2375. https://doi.org/10.1007/s00122-020-03604-1

81) Luk, K. F. Process development for separation and purification of schisandrin B stereoisomer

from fructus schisandrae The Hong Kong University of Science and Technology Library]. http://dx.doi.org/10.14711/thesis-b987389

82) Maccaferri, M., Ricci, A., Salvi, S., Milner, S. G., Noli, E., Martelli, P. L., Casadio, R., Akhunov, E., Scalabrin, S., Vendramin, V., Ammar, K., Blanco, A., Desiderio, F., Distelfeld, A., Dubcovsky, J., Fahima, T., Faris, J., Korol, A., Massi, A., . . . Tuberosa, R. (2014). A high-density, <scp>SNP</scp>-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. Plant Biotechnology Journal, 13(5), 648-663. https://doi.org/10.1111/pbi.12288

83) Madlung, A., Masuelli, R. W., Watson, B., Reynolds, S. H., Davison, J., & Comai, L. (2002). Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids. Plant physiology, 129(2), 733-746. https://doi.org/10.1104/pp.003095

84) Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., Martienssen, R., & Comai, L. (2004). Genomic changes in synthetic Arabidopsis polyploids. The Plant Journal, 41(2), 221-230. https://doi.org/10.1111/j.1365-313x.2004.02297.x

85) Manjunath, K. K., Bs, H., Nm, A., & Ha, B. (2023). Breeding for mulberry improvement: A review. The Pharma Innovation, 12(2), 477-481. https://doi.org/10.22271/tpi.2023.v12.i2f.18461

86) Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics, 9(1), 387-402. https://doi.org/10.1146/annurev.genom.9.081307.164359

87) Mason, A. S. (2014). SSR Genotyping. In Methods in Molecular Biology (pp. 77-89): Springer New York.

88) Mathew, F., Varghese, B., Joseph, D., Mathews, M. M., Sunny, B., & George, J. (2019). An Appraisal of Pharmacological actions of Morus indica: The Indian mulberry with a detailed investigation on its Anti-diabetic potential. Research Journal of Pharmacy and Technology, 12(8), 3654. https://doi.org/10.5958/0974-360x.2019.00623.1

89) Moffat, J. W. (2006). Scalar–tensor–vector gravity theory. Journal of Cosmology and Astroparticle Physics, 2006(03), 004-004. https://doi.org/10.1088/1475-7516/2006/03/004

90) Mushtaq, R., Qadiri, B., Lone, F. A., Raja, T. A., Singh, H., Pervez, A., & Sharma, R. (2023). Role of Sericulture in Achieving Sustainable Development Goals. Problemy Ekorozwoju, 18(1), 199-206. https://doi.org/10.35784/pe.2023.1.21

91) Negro, C., Aprile, A., De Bellis, L., & Miceli, A. (2019). Nutraceutical Properties of Mulberries Grown in Southern Italy (Apulia). Antioxidants (Basel, Switzerland), 8(7), 223.

https://doi.org/10.3390/antiox8070223

92) Newell, W. R., Mott, R., Beck, S., & Lehrach, H. (1995). Construction of Genetic Maps Using Distance Geometry. Genomics, 30(1), 59-70. https://doi.org/10.1006/geno.1995.0009

93) Nicolas, S. D., Leflon, M., Monod, H., Eber, F., Coriton, O., Huteau, V., Chèvre, A.-M., & Jenczewski, E. (2009). Genetic regulation of meiotic cross-overs between related genomes in Brassica napus haploids and hybrids. The Plant cell, 21(2), 373-385. https://doi.org/10.1105/tpc.108.062273

94) Ou, T., Zhang, M., Gao, H., Wang, F., Xu, W., Liu, X., Wang, L., Wang, R., & Xie, J. (2023). Study on the Potential for Stimulating Mulberry Growth and Drought Tolerance of Plant Growth-Promoting Fungi. International journal of molecular sciences, 24(4), 4090. https://doi.org/10.3390/ijms24044090

95) Pada Bag, M. (2021). Use of Sericulture Byproduct as Feed for Tilapia – Oreochromis niloticus (Linn.). Bioscience Biotechnology Research Communications, 14(2), 622-627. https://doi.org/10.21786/bbrc/14.2.28

96) Park, K. M., You, J. S., Lee, H. Y., Baek, N. I., & Hwang, J. K. (2003). Kuwanon G: an antibacterial agent from the root bark of Morus alba against oral pathogens. Journal of Ethnopharmacology, 84(2-3), 181-185. https://doi.org/10.1016/s0378-8741(02)00318-5

97) Park, K. M., You, J. S., Lee, H. Y., Baek, N. I., & Hwang, J. K. (2003). Kuwanon G: an antibacterial agent from the root bark of Morus alba against oral pathogens. Journal of Ethnopharmacology, 84(2-3), 181-185. https://doi.org/10.1016/s0378-8741(02)00318-5

98) Pérez-Gregorio, M. R., Regueiro, J., Alonso-González, E., Pastrana-Castro, L. M., & Simal-Gándara, J. (2011). Influence of alcoholic fermentation process on antioxidant activity and phenolic levels from mulberries (Morus nigra L.). LWT - Food Science and Technology, 44(8), 1793-1801. https://doi.org/10.1016/j.lwt.2011.03.007

99) Pinto, M. V., H S, P., M S, R., R, T., & Naik, V. G. (2018). Association mapping of quantitative resistance to charcoal root rot in mulberry germplasm. PloS one, 13(7), e0200099-e0200099. https://doi.org/10.1371/journal.pone.0200099

100) Pinto, M. V., H S, P., M S, R., R, T., & Naik, V. G. (2018). Association mapping of quantitative resistance to charcoal root rot in mulberry germplasm. PloS one, 13(7), e0200099-e0200099. https://doi.org/10.1371/journal.pone.0200099

101) Pires, J. C., Lim, K. Y., Kovarík, A., Matyásek, R., Boyd, A., Leitch, A. R., Leitch, I. J., Bennett, M. D., Soltis, P. S., & Soltis, D. E. (2004). Molecular cytogenetic analysis of recently evolvedTragopogon(Asteraceae) allopolyploids reveal a karyotype that is additive of the diploid progenitors. American Journal of Botany, 91(7), 1022-1035.

https://doi.org/10.3732/ajb.91.7.1022

102) Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J. R., Sonthirod, C., Sangsrakru, D., Tragoonrung, S., & Tangphatsornruang, S. (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). Genomics, 105(5-6), 288-295. https://doi.org/10.1016/j.ygeno.2015.02.002

103) Porceddu, A., Albertini, E., Barcaccia, G., Falistocco, E., & Falcinelli, M. (2002). Linkage mapping in apomictic and sexual Kentucky bluegrass (Poa pratensis L.) genotypes using a two way pseudo-testcross strategy based on AFLP and SAMPL markers. Theoretical and Applied Genetics, 104(2), 273-280. https://doi.org/10.1007/s001220100659

104) Radojkovic, M. Mulberry extracts (Morus spp., Morasceae), composition, activity and application National Library of Serbia]. http://dx.doi.org/10.2298/ns20121109radojkovic

105) Ramsey, J., & Schemske, D. W. (2002). Neopolyploidy in Flowering Plants. Annual Review of Ecology and Systematics, 33(1), 589-639. https://doi.org/10.1146/annurev.ecolsys.33.010802.150437

106) Ray, P., Gogoi, S. N., Bandyopadhyay, S., Jena, R. K., Ramachandran, S., Sharma, G. K., Sah, K. D., Singh, S. K., & Ray, S. K. (2021). Soil-cum-nutrient management package for achieving high yield and quality of mulberry (Morus indica L.) leaf in acid soil of North Eastern India. Journal of Plant Nutrition, 44(17), 2501-2513. https://doi.org/10.1080/01904167.2021.1921196

107) Ray, P., Gogoi, S. N., Bandyopadhyay, S., Jena, R. K., Ramachandran, S., Sharma, G. K., Sah, K. D., Singh, S. K., & Ray, S. K. (2021). Soil-cum-nutrient management package for achieving high yield and quality of mulberry (Morus indica L.) leaf in acid soil of North Eastern India. Journal of Plant Nutrition, 44(17), 2501-2513. https://doi.org/10.1080/01904167.2021.1921196

108) Ray, P., Gogoi, S. N., Bandyopadhyay, S., Jena, R. K., Ramachandran, S., Sharma, G. K., Sah, K. D., Singh, S. K., & Ray, S. K. (2021). Soil-cum-nutrient management package for achieving high yield and quality of mulberry (Morus indica L.) leaf in acid soil of North Eastern India. Journal of Plant Nutrition, 44(17), 2501-2513. https://doi.org/10.1080/01904167.2021.1921196

109) Reyna-López, G. E., Simpson, J., & Ruiz-Herrera, J. (1997). Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Molecular and General Genetics MGG, 253(6), 703-710. https://doi.org/10.1007/s004380050374

110) Ronin, Y., Mester, D., Minkov, D., Belotserkovski, R., Jackson, B. N., Schnable, P. S., Aluru, S., & Korol, A. (2012). Two-phase analysis in consensus genetic mapping. G3 (Bethesda, Md.), 2(5), 537-549. https://doi.org/10.1534/g3.112.002428

111) Rukmangada, M. S., Sumathy, R., Sivaprasad, V., & Naik, V. G. (2018). Genome-wide identification and characterization of growth-regulating factors in mulberry (Morus spp.). Trees, 32(6), 1695-1705. https://doi.org/10.1007/s00468-018-1744-6

112) Saini, P., Rohela, G. K., Kumar, J. S., Shabnam, A. A., & Kumar, A. (2023). Cultivation, Utilization, and Economic Benefits of Mulberry. In Compendium of Plant Genomes (pp. 13-56): Springer International Publishing.

113) Saini, P., Rohela, G. K., Kumar, J. S., Shabnam, A. A., & Kumar, A. (2023). Cultivation, Utilization, and Economic Benefits of Mulberry. In Compendium of Plant Genomes (pp. 13-56): Springer International Publishing.

114) Sajeevan, R. S., Nataraja, K. N., Shivashankara, K. S., Pallavi, N., Gurumurthy, D. S., & Shivanna, M. B. (2017). Expression of Arabidopsis SHN1 in Indian Mulberry (Morus indica L.) Increases Leaf Surface Wax Content and Reduces Post-harvest Water Loss. Frontiers in plant science, 8, 418-418. https://doi.org/10.3389/fpls.2017.00418

115) Sajeevan, R. S., Nataraja, K. N., Shivashankara, K. S., Pallavi, N., Gurumurthy, D. S., & Shivanna, M. B. (2017). Expression of Arabidopsis SHN1 in Indian Mulberry (Morus indica L.) Increases Leaf Surface Wax Content and Reduces Post-harvest Water Loss. Frontiers in plant science, 8, 418-418. https://doi.org/10.3389/fpls.2017.00418

116) Salmon, A., Ainouche, M. L., & Wendel, J. F. (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Molecular Ecology, 14(4), 1163-1175. https://doi.org/10.1111/j.1365-294x.2005.02488.x

117) San, B., & Yildirim, A. N. (2010). Phenolic, alpha-tocopherol, beta-carotene and fatty acid composition of four promising jujube (Ziziphus jujuba Miller) selections. Journal of Food Composition and Analysis, 23(7), 706-710. https://doi.org/10.1016/j.jfca.2010.02.008

118) Sarkar, T., Ravindra, K. N., Doss, S. G., Kumar, P. M. P., & Tewary, P. (2021). In vitro regeneration of mulberry plants from seedling explants of Morus indica cv. G4 through direct organogenesis. Trees, 36(1), 113-125. https://doi.org/10.1007/s00468-021-02186-9

119) Sastry, G. R. K. (1984). Molecular biology of plant tumours. Physiological Plant Pathology, 24(1), 129-130. https://doi.org/10.1016/0048-4059(84)90080-8

120) Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. Human Molecular Genetics, 19(R2), R227-R240. https://doi.org/10.1093/hmg/ddq416

121) Shreelakshmi, S. V., Nazareth, M. S., Kumar, S. S., Giridhar, P., Prashanth, K. V. H., & Shetty, N. P. (2021). Physicochemical Composition and Characterization of Bioactive Compounds of Mulberry (Morus indica L.) Fruit During Ontogeny. Plant Foods for Human Nutrition, 76(3), 304-310. https://doi.org/10.1007/s11130-021-00909-4

122) Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S. P., Burns, P., Davis, T. M., Slovin, J. P., Bassil, N., Hellens, R. P., Evans, C., Harkins, T., Kodira, C., Desany, B., . . . Folta, K. M. (2011). The genome of woodland strawberry (Fragaria vesca). Nature genetics, 43(2), 109-116. https://doi.org/10.1038/ng.740

123) Somers, D. J., Isaac, P., & Edwards, K. (2004). A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.). Theoretical and Applied Genetics, 109(6), 1105-1114. https://doi.org/10.1007/s00122-004-1740-7

124) Song, K., Lu, P., Tang, K., & Osborn, T. C. (1995). Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proceedings of the National Academy of Sciences of the United States of America, 92(17), 7719-7723. https://doi.org/10.1073/pnas.92.17.7719

125) Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. The Plant Journal, 3(5), 739-744. https://doi.org/10.1046/j.1365-313x.1993.03050739.x

126) Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., Jiang, C., Guan, N., Ma, C., Zeng, H., Xu, C., Song, J., Huang, L., Wang, C., Shi, J., Wang, R., Zheng, X., Lu, C., Wang, X., & Zheng, H. (2013). SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PloS one, 8(3), e58700-e58700. https://doi.org/10.1371/journal.pone.0058700

127) Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Huneau, C., Belcram, H., Coriton, O., Manzanares-Dauleux, M. J., Delourme, R., King, G. J., Chalhoub, B., Jenczewski, E., & Chèvre, A. M. (2010). The first meiosis of resynthesized Brassica napus, a genome blender. New Phytologist, 186(1), 102-112. https://doi.org/10.1111/j.1469-8137.2010.03182.x

128) Tate, J. A., Soltis, D. E., & Soltis, P. S. (2005). Polyploidy in Plants. In The Evolution of the Genome (pp. 371-426): Elsevier.

129) Teh, C.-K., Ong, A.-L., Kwong, Q.-B., Apparow, S., Chew, F.-T., Mayes, S., Mohamed, M., Appleton, D., & Kulaveerasingam, H. (2016). Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Scientific reports, 6, 19075-19075. https://doi.org/10.1038/srep19075

130) Thabti, I., Elfalleh, W., Tlili, N., Ziadi, M., Campos, M. G., & Ferchichi, A. (2013). Phenols, Flavonoids, and Antioxidant and Antibacterial Activity of Leaves and Stem Bark ofMorusSpecies. International Journal of Food Properties, 17(4), 842-854. https://doi.org/10.1080/10942912.2012.660722

131) Thaipitakwong, T., Numhom, S., & Aramwit, P. (2018). Mulberry leaves and their potential effects against cardiometabolic risks: a review of chemical compositions, biological properties and clinical efficacy. Pharmaceutical biology, 56(1), 109-118. https://doi.org/10.1080/13880209.2018.1424210

132) Tikader, A., & Thangavelu, K. (2006). Grafting performance in some Mulberry germplasm collected through exploration. Indian Journal of Forestry, 29(1), 69-71. https://doi.org/10.54207/bsmps1000-2006-29xr38

133) Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., Hine, E. E., Althoff, R., Arbogast, T. S., Tallon, L. J., Vigouroux, M., Trick, M., & Bancroft, I. (2006). Comparative genomics of Brassica oleracea and Arabidopsis thaliana reveal gene loss, fragmentation, and dispersal after polyploidy. The Plant cell, 18(6), 1348-1359. https://doi.org/10.1105/tpc.106.041665

134) Van, O., W., J., & Jansen, J. (2013). Preface. In Genetic Mapping in Experimental Populations (pp. xi-xii): Cambridge University Press.

135) Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., Haudenschild, C. D., Moore, S. S., Warren, W. C., & Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods, 5(3), 247-252. https://doi.org/10.1038/nmeth.1185

136) Vijayan, K., Nair, C. V., & Chatterjee, S. N. (2005). Molecular characterization of mulberry genetic resources indigenous to India. Genetic Resources and Crop Evolution, 52(1), 77-86. https://doi.org/10.1007/s10722-005-0288-y

137) Vijayan, K., Ravikumar, G., & Tikader, A. (2018). Mulberry (Morus spp.) Breeding for Higher Fruit Production. In Advances in Plant Breeding Strategies: Fruits (pp. 89-130): Springer International Publishing.

138) Vijayan, K., Srivastava, P. P., Raju, P. J., & Saratchandra, B. (2012). Breeding for higher productivity in mulberry. Czech Journal of Genetics and Plant Breeding, 48(4), 147-156. https://doi.org/10.17221/162/2011-cjgpb

139) Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing, C., Lillemo, M., Mather, D.,

. . . Akhunov, E. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant biotechnology journal, 12(6), 787-796. https://doi.org/10.1111/pbi.12183

140) Wang, Y., Jin, W., Che, Y., Huang, D., Wang, J., Zhao, M., & Sun, G. (2019). Atmospheric Nitrogen Dioxide Improves Photosynthesis in Mulberry Leaves via Effective Utilization of Excess Absorbed Light Energy. Forests, 10(4), 312. https://doi.org/10.3390/f10040312

141) Wang, Y., Jin, W., Che, Y., Huang, D., Wang, J., Zhao, M., & Sun, G. (2019). Atmospheric Nitrogen Dioxide Improves Photosynthesis in Mulberry Leaves via Effective Utilization of Excess Absorbed Light Energy. Forests, 10(4), 312. https://doi.org/10.3390/f10040312

142) Wang, Z., Hu, H., Jiang, X., Tao, Y., Lin, Y., Wu, F., Hou, S., Liu, S., Li, C., Chen, G., & Liu, Y. (2020). Identification and Validation of a Novel Major Quantitative Trait Locus for Plant Height in Common Wheat (Triticum aestivum L.). Frontiers in genetics, 11, 602495-602495. https://doi.org/10.3389/fgene.2020.602495

143) Wen, P., Hu, T.-G., Linhardt, R. J., Liao, S.-T., Wu, H., & Zou, Y.-X. (2019). Mulberry: A review of bioactive compounds and advanced processing technology. Trends in Food Science & Technology, 83, 138-158. https://doi.org/10.1016/j.tifs.2018.11.017

144) Whelan, J. (2008). The health implications of changing linoleic acid intakes. Prostaglandins, Leukotrienes and Essential Fatty Acids, 79(3-5), 165-167. https://doi.org/10.1016/j.plefa.2008.09.013

145) Wingen, L. U., West, C., Leverington-Waite, M., Collier, S., Orford, S., Goram, R., Yang, C.-Y., King, J., Allen, A. M., Burridge, A., Edwards, K. J., & Griffiths, S. (2017). Wheat Landrace Genome Diversity. Genetics, 205(4), 1657-1676. https://doi.org/10.1534/genetics.116.194688

146) Won-Jin, C., Ji-Ho, C., Won-Taek, C., Seung-Hee, K., Jong-Seok, K., Shi-Jong, L., & Tae-Kyung, Y. Hydrogen effect on 670 nm AlGaInP visible laser during high temperature operation Proceedings of IEEE 14th International Semiconductor Laser Conference, http://dx.doi.org/10.1109/islc.1994.519152

147) Xie, J.-H., Jin, M.-L., Morris, G. A., Zha, X.-Q., Chen, H.-Q., Yi, Y., Li, J.-E., Wang, Z.-J., Gao, J., Nie, S.-P., Shang, P., & Xie, M.-Y. (2015). Advances on Bioactive Polysaccharides from Medicinal Plants. Critical Reviews in Food Science and Nutrition, 56(sup1), S60-S84. https://doi.org/10.1080/10408398.2015.1069255

148) Yang, X., Yang, L., & Zheng, H. (2010). Hypolipidemic and antioxidant effects of

mulberry (Morus alba L.) fruit in hyperlipidaemia rats. Food and Chemical Toxicology, 48(8-9), 2374-2379. https://doi.org/10.1016/j.fct.2010.05.074

149) Yonghui, W., Close, T. J., & Lonardi, S. (2011). Accurate Construction of Consensus Genetic Maps via Integer Linear Programming. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8(2), 381-394. https://doi.org/10.1109/tcbb.2010.35

150) Yuan, Q., & Zhao, L. (2017). The Mulberry (Morus alba L.) Fruit—A Review of Characteristic Components and Health Benefits. Journal of Agricultural and Food Chemistry, 65(48), 10383-10394. https://doi.org/10.1021/acs.jafc.7b03614

151) Zhang, Y., Wang, L., Xin, H., Li, D., Ma, C., Ding, X., Hong, W., & Zhang, X. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biology, 13, 141-141. https://doi.org/10.1186/1471-2229-13-141

152) Zhang, Z., Shang, H., Shi, Y., Huang, L., Li, J., Ge, Q., Gong, J., Liu, A., Chen, T., Wang, D., Wang, Y., Palanga, K. K., Muhammad, J., Li, W., Lu, Q., Deng, X., Tan, Y., Song, W., Cai, J., . . . Yuan, Y. (2016). Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (Gossypium hirsutum.). BMC Plant Biology, 16, 79-79. https://doi.org/10.1186/s12870-016-0741-4

153) Zhang, Z., Shang, H., Shi, Y., Huang, L., Li, J., Ge, Q., Gong, J., Liu, A., Chen, T., Wang, D., Wang, Y., Palanga, K. K., Muhammad, J., Li, W., Lu, Q., Deng, X., Tan, Y., Song, W., Cai, J., . . . Yuan, Y. (2016). Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (Gossypium hirsutum.). BMC Plant Biology, 16, 79-79. https://doi.org/10.1186/s12870-016-0741-4

154) Zhao, L., Wu, M., Xiao, C., Yang, L., Zhou, L., Gao, N., Li, Z., Chen, J., Chen, J., Liu, J., Qin, H., & Zhao, J. (2015). Discovery of an intrinsic tenase complex inhibitor: Pure nonasaccharide from fucosylated glycosaminoglycan. Proceedings of the National Academy of Sciences of the United States of America, 112(27), 8284-8289. https://doi.org/10.1073/pnas.1504229112

155) Zhao, M., Shu, G., Hu, Y., Cao, G., & Wang, Y. (2023). Pattern and variation in simple sequence repeat (SSR) at different genomic regions and its implications to maize evolution and breeding. BMC genomics, 24(1), 136-136. https://doi.org/10.1186/s12864-023-09156-0