CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Recognizing and Assessing *Sterile Apetala* genes in *Gossypium hirsutum* for their Putative Role in Floral Development: A Bioinformatics Approach

by

Saman Shafique

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2025

I would like to express my deepest gratitude to my parents, whose unwavering love and support have been the foundation of my journey. To my mother, thank you for your endless patience, unconditional love, and constant prayers that have lifted me during times of doubt. Your nurturing spirit and quiet strength have been my greatest comfort. To my father, your wisdom, encouragement, and belief in my abilities have inspired me to pursue my goals with determination and confidence. Your guidance has shaped not only this thesis but also the person I have become. I am truly blessed to have you both by my side, and this achievement is as much yours as it is mine.

CERTIFICATE OF APPROVAL

Recognizing and Assessing *Sterile Apetala* genes in *Gossypium hirsutum* for their Putative Role in Floral Development: A Bioinformatics Approach

by

Saman Shafique

(MBS241007)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. M. Zeeshan Qureshi	MIU, AJK
(b)	Internal Examiner	Dr. Sahar Afzal	CUST, Islamabad
(c)	Supervisor	Dr. Sami Ullah Jan	CUST, Islamabad

Dr. Sami Ullah Jan
Thesis Supervisor
September, 2025

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. & Biosciences
September, 2025

Dr. Sahar Fazal
Dean
Faculty of Health & Life Sciences
September, 2025

Author's Declaration

I, **Saman Shafique** hereby state that my MS thesis titled " **Recognizing and Assessing *Sterile Apetala* genes in *Gossypium hirsutum* for their Putative Role in Floral Development: A Bioinformatics Approach**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Saman Shafique)

Registration No: MBS241007

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled " **Recognizing and Assessing *Sterile Apetala* genes in *Gossypium hirsutum* for their Putative Role in Floral Development: A Bioinformatics Approach**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Saman Shafique)
Registration No: MBS241007

Acknowledgement

Alhamdulillah! All praise and gratitude belong to Allah, the Lord of the worlds. I would like to express my heartfelt gratitude to my supervisor, Dr. Sami Ullah Jan, for his invaluable guidance throughout my thesis journey. I am deeply thankful to my family for their unwavering love and support. Additionally, I am grateful to my friend for her generous help and guidance, which have been instrumental in my progress. Thank you all for being an essential part of this journey.

I also acknowledge the support of Dean **Dr. Sahar Fazal** and **Dr. Syed Marriam Bakhtiar** Head of Department, for their leadership and for providing the excellent research facilities that enabled this bioinformatics study.

I extend my heartfelt thanks to **CUST, Islamabad** for providing me with the necessary resources and support. I am also grateful to my professors for their constructive suggestions and motivation.

Special thanks to my family for their unwavering support, understanding, and encouragement during challenging times. Their belief in me kept me motivated.

Lastly, I acknowledge all the individuals who directly or indirectly contributed to this endeavor. This accomplishment would not have been possible without their help.

(Saman Shafique)

Abstract

The flower serves as a crucial organ for both reproduction and yield in all plants. Floral development plays a pivotal role in determining reproductive success and agricultural productivity, particularly in economically important crops like *Gossypium hirsutum* (upland cotton), a globally significant fiber-producing plant. In cotton, flowers are especially vital as they directly contribute to its status as a cash crop. However, various abiotic stresses significantly impact floral development, leading to substantial economic losses. To address this challenge, researchers employ various techniques to enhance floral development in plants. Among these, genetic engineering of floral development-associated genes has proven particularly effective. One such gene is *Sterile Apetala* (*SAP*), which regulates floral organ identity and stress responses, making it a promising target for improving cotton yield and stress resilience. This study utilized a bioinformatics approach to identify and characterize *SAP* genes in cotton, with particular focus on their potential roles in floral development and abiotic stress tolerance. Two *SAP* genes (*GhSAP1* and *GhSAP2*) were identified in cotton, featuring conserved F-Box and WD40 domains critical for function. Protein interactions linked them to stress response pathways, while promoter analysis revealed regulatory elements tied to hormonal and stress adaptation. Phylogenetic analysis showed evolutionary divergence, suggesting specialized roles in cotton species. The research highlights the promising role of *SAP* genes in optimizing floral development and bolstering stress tolerance in cotton. By uncovering the genetic pathways associated with these traits, this research establishes a framework for precision breeding approaches designed to boost productivity and sustainable cotton farming. Further studies should prioritize experimental validation to fully exploit the agronomic benefits of *SAP* genes in enhancing crop performance.

Keywords: *Sterila Apetala (SAP)*, Flowering, Floral Development, *Gossypium hirsutum*, Cotton, Abiotic Stresses, Genome-Wide Analysis.

Contents

Author's Declaration iv

Plagiarism Undertaking v

Acknowledgement vi

Abstract vii

List of Figures xi

List of Tables xii

Abbreviations xiii

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Problem Statement . . . 5
- 1.3 Aim . . . 5
- 1.4 Objectives . . . 5
- 1.5 Scope of Study . . . 5
- 1.6 Impact on Society . . . 6

2 Literature Review 7
- 2.1 Taxonomy and Economic Importance of Cotton . . . 7
  - 2.1.1 Cotton Production in Pakistan . . . 11
  - 2.1.2 Import of Cotton in Pakistan . . . 12
  - 2.1.3 World Wide Cotton Trade . . . 13
  - 2.1.4 Industrial Use of Cotton . . . 14
- 2.2 Factors Contributing to Low Cotton Yield in the Country . . . 16
  - 2.2.1 Cultivator Selection . . . 16
  - 2.2.2 Soil Preparation . . . 16
  - 2.2.3 Seed Rate and Plant Spacing . . . 17
  - 2.2.4 Sowing/Planting Date . . . 18
  - 2.2.5 Irrigation . . . 19
  - 2.2.6 Nutrients . . . 19
  - 2.2.7 Crop Protection . . . 20

2.2.8 Weeds . . . 21
2.2.9 Insects . . . 21
2.2.10 Diseases . . . 22
2.2.11 Modern Technology . . . 23
2.3 Genomic Features . . . 23
2.4 Floral Morphology and Development . . . 24
2.5 Mechanism of Cotton Resistance to Abiotic Stress . . . 25
2.5.1 Effects of Non-Living Stress on Cotton Growth and Development . . . 27
2.5.2 Effects of Abiotic Stress on Cotton Growth and Development 27
2.5.3 Effects of Salt and Alkali on Cotton Growth and Development 28
2.5.4 Influence of Temperature Change on Cotton Growth . . . 30
2.5.5 Effects of Soil Nutrient Stress on Cotton Growth and Development . . . 32
2.6 Transcription Factors and Plant Growth . . . 33
2.6.1 Floral Development . . . 35
2.6.2 Leaf Development . . . 36
2.6.3 Root Development . . . 37
2.6.4 Transcription Factors and Stress Responses in Plants . . . 37
2.6.5 Major TF Families Involved in Stress Responses . . . 38
2.6.6 Role of NAC Transcription Factors . . . 38
2.6.7 MYB Transcription Factors and Stress Regulation . . . 39
2.6.8 WRKY TFs in Stress Defense . . . 39
2.6.9 Heat Shock Factors (HSFs) . . . 39
2.6.10 *AP2/ERF* Family in Temperature Stress . . . 39
2.6.11 Transcription Factors and Secondary Metabolism . . . 42
2.7 Molecular Basis of Floral Development in Plants . . . 44
2.7.1 ABCDE Model of Flower Development . . . 44
2.7.2 Key Regulatory Genes in Flower Development . . . 46
2.7.3 Role of MADS-box Transcription Factors . . . 47
2.8 The *Sterile Apetala* (*SAP*) Gene Family . . . 48
2.8.1 Historical Discovery and Characterization . . . 48
2.8.2 Structure and Domains of *SAP* Genes . . . 49
2.8.2.1 MADS Domain (M) . . . 49
2.8.2.2 Intervening Domain (I) . . . 49
2.8.2.3 Keratin-like Domain (K) . . . 49
2.8.2.4 C-terminal Domain (C) . . . 50
2.8.3 Functional Role in Model Plants . . . 50
2.8.4 *SAP* Orthologs and Paralogs in Other Species . . . 51
2.9 *Sterile Apetala* Genes In *Gossypium* Species . . . 52
2.9.1 Identification of *SAP* Genes in *Gossypium* . . . 52
2.9.2 Evolutionary Divergence and Gene Duplication Event . . . 53
2.9.3 Reported Functional Studies . . . 54
2.10 Bioinformatics Approaches and Pipelines . . . 55
2.10.1 Databases and Genomic Resources for Cotton . . . 55

2.10.2 Tools for Gene Structure Prediction . . . 56
2.10.3 Phylogenomic Methods . . . 56
2.10.4 Functional Annotation Tools . . . 57
2.10.5 Integrative Approaches . . . 58

3 Methodology **60**
3.1 Screening of Genome and Transcription Factors Databases . . . 61
3.2 GhSAPs Protein Sequences Retrieval . . . 61
3.3 Removal of Redundant Sequences of *GhSAPs* . . . 61
3.4 Allocation of Lab IDs and Establishment of Sequence Similarity of *GhSAP1* and *GhSAP2* Protein Sequences . . . 62
3.5 Physicochemical Properties and Conserved Domains in *AtSAP* and *GhSAP1* and *GhSAP2* . . . 62
3.6 Motif Composition in *GhSAPs* Gene Family . . . 62
3.7 Phylogenetic Analysis of *GhSAPs* . . . 63
3.8 Gene structure and Motif Composition in *GhSAPs* Gene Family . . . 63
3.9 Protein-Protein Interaction of *GhSAPs* . . . 63
3.10 Promoter Analysis . . . 64

4 Results **65**
4.1 Genome - Wide Identification and Analysis of *SAP* Gene in *Gossypium hirsutum* . . . 65
4.2 Physicochemical Properties and Conserved Domains of *Gossypium hirsutum* . . . 66
4.3 Gene Structure Determination . . . 67
4.4 Displays the Result of Gene Structure Determination . . . 68
4.5 Consensus Motifs Composition in ATSAPs and GhSAPs . . . 71
4.6 Logos of Identified Motifs . . . 72
4.7 Sequence Alignment and Phylogenetic Relationship of *GhSAP* Gene Family . . . 72
4.8 Promoter Analysis . . . 75
4.8.1 *AtSAP1* . . . 75
4.8.2 *GhSAP1* . . . 75
4.8.3 *GhSAP2* . . . 76

5 Discussion **77**

6 Conclusion **81**
6.1 Future Recommendations . . . 82

Bibliography **83**

List of Figures

2.1 Cotton Plant Anatomy 8
2.2 Top cotton producing countries worldwide 9
2.3 Production and consumption of world's major cotton growing countries 9
2.4 Cotton production in Pakistan (000 bales) 11
2.5 Physiological, biochemical and molecular basis of drought resistance in plants 26
2.6 The complexity of plant response to adversity stress 29
2.7 Representation of class A, B, C, D, and E genes with their respective tissue in model flower 45

3.1 Flowchart of Methodology 60

4.1 Displays the result of Gene Structure Determination 68
4.2 *AtSAP* Protein-Protein Interaction 69
4.3 *GhSAP1* Protein Protein Interaction 70
4.4 *GhSAP2* Protein-Protein Interactions 71
4.5 Consensus motifs of *AtSAP* and *GhSAPs* 72
4.6 Logos of Identified Motifs 72
4.7 Phylogenetic analysis of *GhSAPs* through neighbour joining method using MEGA 74

List of Tables

2.1 Transcription factors identified for their role in plant development [83]. 34
2.2 Exemplification of the contribution of transcription factors in stress challenged plants. 40
2.3 Examples of transcription factors in secondary metabolism. 43

4.1 Sequence similarity among protein sequences of GhSAPs 66
4.2 Conserved Domains of *SAP* gene families of *GhSAP1* and *GhSAP2* 67
4.3 Physical and chemical properties of SAP gene families of *Arabidopsis thaliana* and Gossypium hirsutum 67
4.4 AtSAP Protein Protein Interaction 69
4.5 *GhSAP1* Protein-Protein Interaction 70
4.6 GhSAP2 Protein-Protein Interaction 71
4.7 Plant Growth, Stress, and Phytohormone Responsive Motifs in Genes 76

Abbreviations

AP 1	Apetela 1
AP2/ERF	APETALA2/Ethylene Responsive Factor
AP3	Apetela 3
bHLH	Basic Helix-Loop-Helix
BLAST	Basic Local Alignment search Tool
CLCuV	Cotton Leaf Curl Virus
CUC	Cup-shaped Cotyledon
EGL3	Enhancer of GLABRA3
Expasy	Expert Protein Analysis System
GL3	GLABRA33
GRAVY	Grand Average of Hydropathicity
GRF	Growth regulating factors
GSDS	Gene Structure Display Server
HMMs	Hidden Markov Model
HSFs	Heat Shock Factors
ITOL	Interactive Tree of Life Software v6
LFY	LEAFY
MEME	Multiple Em foe Motif Elication
MYB	Myeloblastosis-related
NCBI	National Centre for Biotechnology Information
PGD	Plant Growth and Development
PR	Phytohormone
PlantTFBD	Plant transcription Factor Database
ROS	Reactive oxygen species

SAP	Sterile Apetela
SOC1	Supressor of overexpression of Constansi
TAIR	The Arabidopsis Information Resource (TAIR)
TFs	Transcription factors
UTRs	Untranslated Regions on mRNA
VIGS	Virus-Induced Gene Silencing
WOX	Wuschel-related HomeBo

Chapter 1

Introduction

1.1 Background

Floral development is an important part of a plant's life cycle, and it affects how well plants can reproduce, as well as how productive and high quality crops are. In flowering plants, getting the right structures in the flower is essential for pollination, fertilization, and making seeds and fruits. The impact of how flowers develop can have big economic effects, especially for crops like *Gossypium hirsutum*, which is known as upland cotton [1]. Cotton is one of the main sources of natural fiber in the world. Growing cotton supports the income of millions of farmers around the globe, which makes studying the genetic and molecular aspects of its growth very important [2]. In plant biology, figuring out how flowers develop has been a major focus for a long time. A group of proteins called MADS-box transcription factors plays a key role in this process. These proteins help determine the shape and structure of different parts of a flower through networks of genetic control that are similar across many plant species. Among these factors, the *STERILE APETALA* (*SAP*) genes stand out because they help define the identity of petals and stamen. If these genes don't work properly, it can cause big changes in how the flower looks, like missing petals, which can make the plant unable to reproduce or reduce its ability to have offspring [3].

In *Gossypium hirsutum*, figuring out how *SAP* genes work is important both in theory and in practice. This crop is very important worldwide for making textiles and other products, so there's a need to keep improving its yield, quality, and ability to handle environmental challenges [4]. Features of the flowers, like their size, shape, fertility, and when they bloom, all affect how many bolls form and how much fiber is produced. By learning about the genetic processes that control how flower parts develop, plant breeders can focus on important genetic pathways. This could help create new cotton varieties that are more efficient at reproducing and produce a more stable yield [5].

Furthermore, the study of *SAP* genes contributes to broader questions in evolutionary biology, such as how floral diversity arises and adapts in different lineages [6]. Comparative analyses of SAP genes across species can reveal the evolutionary pressures shaping floral development and can inform phylogenetic relationships within angiosperms. This is particularly relevant in the *Malvaceae* family, where *Gossypium* species exhibit complex evolutionary histories marked by polyploidization and interspecific hybridization events [7].

Floral development is a highly orchestrated process regulated by a network of genetic and hormonal signals. In flowering plants, specific genes determine the identity and formation of floral organs such as sepals, petals, stamens, and carpels. Among these regulatory genes, *STERILE APETALA* (*SAP*) genes play a crucial role in the transition from vegetative to reproductive growth and in maintaining floral meristem identity [8].

Morphogenesis in higher plants begins with an undifferentiated group of cells known as the meristem. Following floral induction, the cells in the shoot apical meristem undergo a change in identity, transforming into the inflorescence meristem. This meristem exhibits indeterminate growth and produces flower meristems along its sides. In *Arabidopsis*, these flower meristems are arranged in a spiral phyllotactic pattern and, unlike the inflorescence meristem, are determinate [9]. The flower meristems give rise to floral organ primordia in a specific number and arrangement. *Arabidopsis* flowers consist of four sepals in the first whorl, four petals in the second, six stamens in the third, and a central pistil [10].

The pistil, formed by the congenital fusion of two carpels, initiates new primordia that develop into ovules. The formation and identity specification of these primordia depend on a complex network of regulatory genes [11].

Many of these genes have been identified in *Arabidopsis* through molecular genetic techniques. However, despite significant progress and gene discovery, gaps remain in our understanding of how these genes interact during flower development, limiting a complete picture of flower ontogeny [12].

The meristem identity genes *LEAFY* (*LFY*) and *APETALA1* (*AP1*) in *Arabidopsis* play crucial roles in establishing the floral meristem. Mutations in either of these genes can partially transform a floral meristem back into an inflorescence meristem.

For example, strong *AP1* mutants produce highly branched, inflorescence-like flowers, often with abnormal secondary flowers developing in the axils of the first whorl organs *LFY* and *AP1* work together synergistically with other floral meristem identity genes, including *APETALA2* (*AP2*) [13].

The *AP2* gene is central to the network that controls both the establishment and maintenance of floral meristem and organ identity. Beyond its role in meristem identity, *AP2* specifies the identity of organs in the first and second floral whorls. It also regulates the timing and location of expression of the homeotic gene *AGAMOUS* (*AG*) in the perianth whorls and in ovules additionally, defects in the seed coat of *AP2* mutants indicate that *AP2* is essential for proper seed development [14].

The homeotic gene *AG* is multifunctional, controlling at least three key aspects of flower development. First, it defines the identity of stamens and carpels. Second, it is critical for floral meristem determination, as *AP2* mutants show indeterminate floral growth, repeatedly producing floral structures within the flower.

Third, according to the ABC model, AG negatively regulates *AP2* activity in stamens and carpels. However, this regulation does not seem to occur at the

transcriptional level, since *AP2* mRNA is present alongside AG expression in the inner two floral whorls [15].

Many genes that play important roles in floral meristem determination and flower organ formation also contribute to ovule and seed development. In some cases, their functions may be hidden due to redundancy with other genes [16]. For instance, the *SUPERMAN* (*SUP*) gene in *Arabidopsis* functions at multiple stages of flower development. It defines the boundary between the organs of the third and fourth floral whorls and also regulates the growth of the integuments, which are the cell layers surrounding the nucleus of the ovule. Another example is the *AINTEGUMENTA* (*ANT*) gene, which is involved in the development of both floral organs and ovules [17].

Originally identified in *Arabidopsis thaliana*, the *SAP* gene encodes a putative transcriptional repressor involved in the control of flowering time, organ identity, and floral determinacy [18]. Loss-of-function mutations in *SAP* often result in abnormal floral structures and reduced fertility, indicating its critical role in reproductive development. Although much of the foundational work on *SAP* genes has been conducted in model plants, their functional roles in economically important crops like *Gossypium hirsutum* remain largely unexplored [19].

In this study, we focus on the identification and analysis of Sterile Apetala (*SAP*) genes in *Gossypium hirsutum*, which are hypothesized to play crucial roles in floral and ovule development, similar to their counterparts in *Arabidopsis* [20]. Drawing from previous research on the *Arabidopsis* sap mutant, where the *SAP* gene was cloned and found to likely function as a transcription regulator, we explore the potential regulatory functions of *SAP* genes in cotton. In *Arabidopsis*, *SAP* works alongside the *AGAMOUS* (*AG*) gene to maintain floral identity, negatively regulating AG expression in the inflorescence meristem and outer floral whorls [21]. Additionally, *SAP* is essential for initiating meiotic divisions during ovule development. Through bioinformatics analyses, including gene sequence characterization, expression profiling, and comparative genomics, this study aims to uncover the role of *SAP* genes in cotton flower morphogenesis. Understanding these regulatory mechanisms will provide valuable insights into the genetic control of floral

development in *Gossypium hirsutum*, which could have important implications for improving fertility and yield in cotton breeding programs [22].

1.2 Problem Statement

Cotton production is severely affected by abiotic stresses, which disrupt key processes like floral development. The *SAP* genes are known to regulate flower formation and stress responses in various plants. However, it is still unknown whether *SAP* exist in cotton and whether it is functionally conserved in cotton.

1.3 Aim

The main aim of this study is to identify and characterize *SAP* gene family in *Gossypium hirsutum* (Upland Cotton) and assess their potentials in enhancing floral development and stress tolerance.

1.4 Objectives

a) To identify *SAP* family in Cotton.

b) To conduct a comprehensive comparison with model plant for assessing the role of *SAP* gene in floral development.

c) To assess the evolutionary relationship of *SAP* gene family.

1.5 Scope of Study

The scope of this study, titled "Recognizing and Assessing *Sterile Apetala* Genes in *Gossypium hirsutum* for Their Putative Role in Floral Development: A Bioinformatics Approach," focuses on the identification and characterization of *Sterile*

Apetala (*SAP*) genes in upland cotton [23]. Using advanced bioinformatics tools, the study aims to analyze the gene sequences, expression patterns, and regulatory networks of SAP genes to understand their potential roles in floral meristem identity and organogenesis [24]. By examining gene expression during various stages of flower development and comparing these findings with model plants such as *Arabidopsis*, the research seeks to uncover the molecular mechanisms that govern flower formation and fertility in cotton. This knowledge is vital for addressing sterility issues and improving reproductive success, which directly impacts fiber yield and quality. Ultimately, the study provides valuable insights that could contribute to cotton breeding programs focused on enhancing floral traits and overcoming challenges related to sterility, thereby supporting efforts to improve cotton productivity and sustainability [25].

1.6 Impact on Society

The study on *Sterile Apetala* genes in *Gossypium hirsutum* holds significant societal impact, particularly in enhancing cotton production, which is a cornerstone of the global textile industry and a vital source of livelihood for millions. By improving our understanding of the genetic mechanisms controlling floral development and fertility in cotton, this research can contribute to developing cotton varieties with better flower formation and higher yields. Increased cotton productivity directly benefits farmers, especially smallholders in developing countries, by improving income stability and reducing vulnerability to environmental stresses [25]. Enhanced yields can lead to greater economic growth in rural areas, supporting employment and improving living standards for farming communities. Furthermore, more efficient and resilient cotton crops can reduce the need for excessive inputs like water and pesticides, promoting sustainable agricultural practices that benefit the environment. Ultimately, advances from this study can help secure the livelihoods of millions dependent on cotton farming, contribute to food and fiber security, and support the socio-economic development of cotton-growing regions worldwide [26].

Chapter 2

Literature Review

2.1 Taxonomy and Economic Importance of Cotton

Upland cotton (*Gossypium hirsutum*), from the *Malvaceae* family, is one of the major economic crops in the world. *Gossypium* is a large genus and has more than 50 species of both diploid and polyploid forms. Of these, *G. hirsutum* contributes about 90% to worldwide cotton production [27].

Its global textile leadership underscores its status as a foundation crop, with millions of farm and textile workers livelihoods at stake in countries around the world [28].

The taxonomic position of *G. hirsutum* is in the *Malvoideae* subfamily, which is evolutionarily related with other fiber-and ornamental-producing species. Through molecular phylogeny we have known some evolutionary relationships of this genus and the speciation of it was complexed and four times of hybridization occurred, which mentioned contributed to the genes diversity [29].

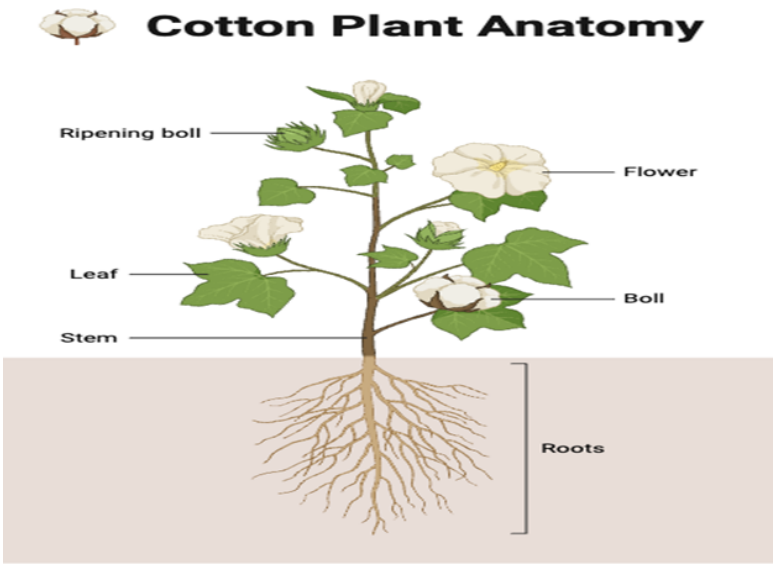

FIGURE 2.1: Cotton Plant Anatomy [29]

From the standpoint of economics, it is highly desirable crop due to its prolificacy, tolerance to various stresses, and the production of long, strong fibers suitable for fabrication into a quality yarn. The crop is not only a raw material for the textile industry but seed oils also derived in by-products of seed meal and edible oil for food processing industry. In developing countries, a significant portion of which is cotton-growing, cotton farming is deeply embedded in rural economies as it acts as a cash crop that supports smallholder farmers [30].

Despite its economic significance, *G. hirsutum* faces persistent challenges, including susceptibility to pests and diseases, vulnerability to climatic extremes, and declining soil fertility due to monoculture practices. Genetic improvement, particularly through molecular breeding and biotechnological interventions, is therefore essential to ensure sustainable cotton production in the face of these challenges [31].

Gossypium hirsutum, commonly known as American cotton or upland cotton, is by far the most widely cultivated cotton species in the world. It accounts for over 90% of global cotton production and is grown on about 95% of the cotton fields across 17 U.S. states, stretching from Virginia to California. In the United States alone, this species makes up roughly 97% of the country's cotton output [32]. The main reasons for its global dominance are its high yield potential and its ability to

adapt to a wide range of growing conditions, which is why it is farmed in more than 80 countries.In contrast, *Gossypium barbadense* better known as Egyptian cotton or Pima cotton is prized for its exceptionally long and fine fibers, which are used to make high-quality textiles. However, this premium cotton variety represents less than 5% of total world cotton production due to its more specialized growing requirements and lower overall yield compared to upland cotton [31].

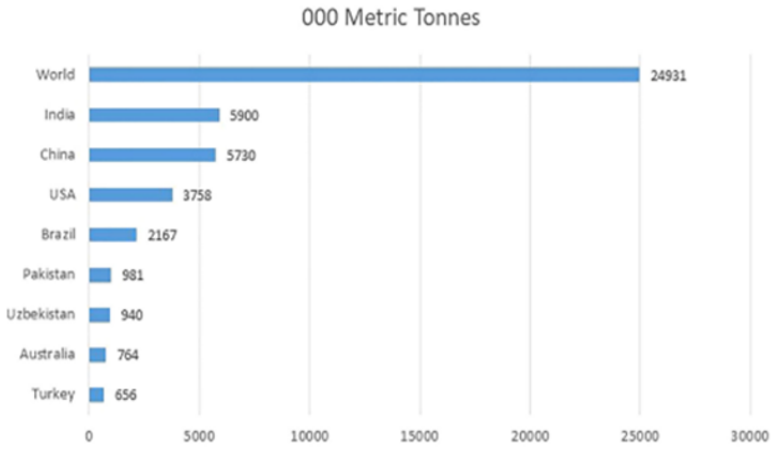

FIGURE 2.2: Top cotton producing countries worldwide [31]

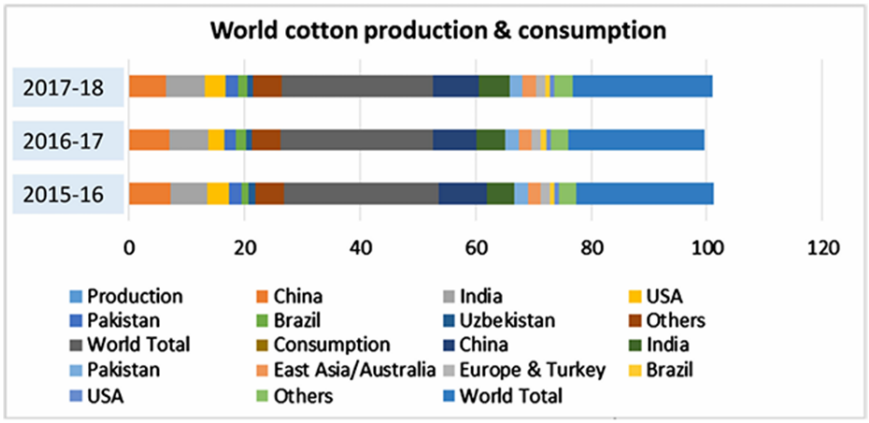

FIGURE 2.3: Production and consumption of world's major cotton growing countries [32]

The cotton we grow today is the product of years of careful breeding, with a strong emphasis on producing high-quality fibers that are easy to harvest and process. However, this intense focus on specific traits has come at a cost. By continually selecting for similar characteristics, the genetic diversity of cultivated cotton has

diminished, particularly in traits like disease resistance and tolerance to drought or salty soils .As these limited genetic resources have been used over time, cotton yields have noticeably declined in recent decades [32].

One promising way to address this issue is to introduce beneficial genes from wild cotton ancestors into modern varieties. This approach can help restore valuable traits and provide solutions to various agricultural challenge. For example, wild species such as *G. tomentosum* and *G. darwinii* are known for their resilience *G. tomentosum* is especially tolerant of salt and drought thanks to its unique agronomic features, while *G. darwinii* offers notable drought resistance, disease resistance, and finer fiber quality [33].

Advances in genomic technologies have played a key role in improving fiber quality and developing new cotton varieties that can withstand both biotic and abiotic stresses. The impact of genetic engineering has been transformative, allowing for the creation of improved cotton genomes. In recent years, a range of "omics" technologies such as transcriptomics, proteomics, and metabolomics have been used to better understand how plants respond to environmental stresses at the molecular level [34].

The growth and development of a typical cotton plant follow a distinct and well-defined pattern. Compared to other major field crops, cotton is known for its complex structure, largely due to its unique indeterminate growth habit and sympodial flowering pattern. Despite this continuous growth, the plant's development can be broken down into several key stages that it progresses through during its life cycle. These stages are generally categorized into five main phases: (i) germination and emergence, (ii) seedling establishment, (iii) leaf area and canopy development, (iv) flowering and boll formation, and (v) maturity.Among these, the flowering and boll development phase stands out as the most critical. During this period, the plant's demand for nutrients and resources rises sharply, making it especially vulnerable to poor management practices and environmental stresses. This stage is therefore crucial for ensuring a healthy and productive cotton crop [33].

Recent projections from the United Nations indicate that the global population could increase by over two billion people from current levels, reaching approximately 9.15 billion by 2050. At the same time, the impacts of climate change are becoming increasingly apparent. Crop production worldwide is facing significant challenges due to both biotic and abiotic stresses, which are being intensified by rapid climate change. To meet the rising demands for food, fiber, clean water, and bioenergy especially in arid and semi-arid regions it is crucial to boost crop production by at least 40%. Therefore, agricultural strategies must focus on developing crops that are resilient to harsh environmental conditions, including stress-tolerant varieties, in order to sustain global agricultural productivity [34].

2.1.1 Cotton Production in Pakistan

Pakistan's production of *Gossypium hirsutum* (upland cotton) has experienced significant fluctuations in recent years. In the 2024-25 season, cotton production sharply declined by about 34%, dropping from 8.39 million bales to approximately 5.52 million bales, with Punjab being the hardest hit region (Punjab's output fell by 36.5%) while Sindh fared somewhat better due to favorable climate and earlier sowing.This decline has been attributed to a worsening water crisis, poor seed germination, pest infestations, and adverse weather conditions, all of which have negatively impacted yields and farmer returns [35].

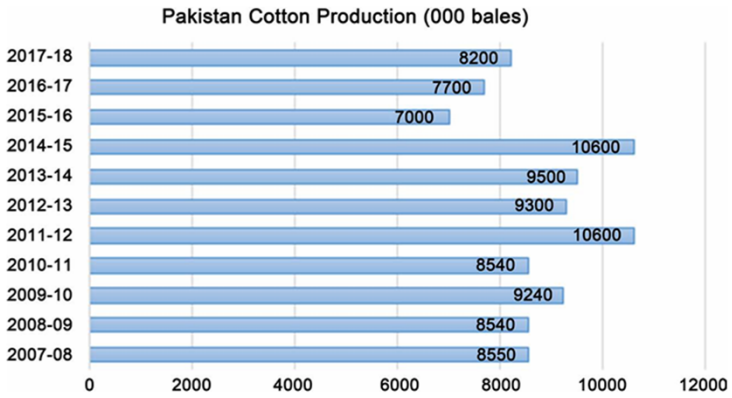

FIGURE 2.4: Cotton production in Pakistan (000 bales) [36].

Cotton cultivation area has also reduced, with only around 66% of the targeted 3.1 million hectares sown during the 2024-25 season, reflecting a 16% decrease compared to the previous year. Punjab and Sindh remain the primary cotton-growing provinces, contributing the majority of production, while smaller contributions come from KPK and Balochistan. The government has recognized these challenges and launched initiatives such as early sowing campaigns, subsidies for farmers, and the introduction of genetically engineered cotton varieties to boost production. The target is to increase annual cotton production to 15 million bales by 2025, aiming to support the textile industry, which is a major pillar of Pakistan's economy [35].

Despite global cotton prices being favorable, local market rates remain weak due to factors like duty-free cotton imports disrupting domestic demand. The ongoing water shortage, particularly in major cotton-producing areas dependent on canal irrigation, poses a serious threat to future crops. Efforts are underway to improve seed quality, irrigation management, and pest control to stabilize and enhance cotton production in the coming years [36].

2.1.2 Import of Cotton in Pakistan

The marketing of *Gossypium hirsutum* in Pakistan faces several challenges but also presents key opportunities for growth. Due to declining domestic cotton production, Pakistan has become heavily reliant on imports, purchasing over 1.5 million bales of cotton and 1.25 million bales of cotton yarn, which has led to an oversupplied market and financial pressure on local farmers and ginners. The presence of an 18% sales tax on domestic cotton further weakens local demand, while duty-free imports increase competition, making it difficult for local producers to compete effectively [32].

To address these issues, the government and industry stakeholders are promoting early cotton cultivation, which has shown promise in optimizing growing seasons, improving yields, and reducing water stress. Punjab's early sowing campaign, supported by subsidies of Rs 25,000 per acre, has successfully increased cotton

acreage, particularly in South Punjab and Bahawalpur, and serves as a model for expanding this approach nationwide. There are also calls for urgent reforms, including tax parity between imported and domestic cotton, investment in heat- and drought-tolerant cotton varieties, modernization of ginning and spinning mills, and improved irrigation infrastructure, especially in regions like Balochistan and Cholistan [37].

Marketing efforts are increasingly focused on stabilizing prices through mechanisms such as minimum support prices and cotton grading standards, which aim to protect farmers' incomes and enhance product quality. Furthermore, expanding cotton cultivation into new areas and promoting organic cotton are part of broader strategies to boost production and market competitiveness [44].

2.1.3 World Wide Cotton Trade

Worldwide cotton trade remains a vital component of the global textile industry, with both exports and imports showing steady growth. For the 2025/26 marketing year, global cotton trade is projected to reach approximately 44 to 45 million bales, marking an increase of about 3.5 million bales compared to the previous year. This growth is primarily driven by rising consumption in emerging textile-producing countries such as Bangladesh, Vietnam, and China, which are among the largest importers. Bangladesh and Vietnam are expected to import around 8.5 million and 8.0 million bales respectively, while China's imports are forecast at about 7 million bales, although this is lower than its peak imports in recent years [38].

On the export side, Brazil has emerged as the world's largest cotton exporter, accounting for roughly 31% of global cotton trade, followed by the United States, which holds about 28-30% share with exports projected near 12.5 to 13.3 million bales. Australia is also a significant exporter, expected to maintain exports around 5.5 million bales. Other regions like West Central Africa contribute smaller but growing shares to global cotton exports [38].

Global cotton production is forecast to be around 25.5 million tons with a slight increase in cultivated area to about 31.5 million hectares. Despite some fluctuations in production due to weather and other factors, consumption is expected to rise steadily, reaching approximately 118 million bales in 2025/26, the highest in five years. This increased demand, coupled with slightly lower production in some countries, is projected to reduce global ending stocks moderately [38].

Cotton prices have experienced some volatility but are expected to gradually stabilize and return to more typical historical levels over the coming years. The overall outlook suggests a steady expansion of cotton trade, supported by growing textile industries in Asia and sustained export capacity from major producers [39].

2.1.4 Industrial Use of Cotton

Cotton, primarily derived from *Gossypium hirsutum*, plays a vital role in various industrial applications, especially within the textile sector. It is widely valued for its softness, breathability, durability, and versatility, making it suitable for producing a broad range of fabrics used in clothing, home furnishings, and industrial products. Textile mills globally manufacture billions of square yards of woven and knitted cotton fabrics annually, which are then used to create apparel such as shirts, jeans, underwear, and outerwear, as well as household items like towels, bedspreads, sheets, and curtains. Cotton dominates markets for towels and washcloths, supplying nearly 100% of these products, and holds a significant share in sheets and pillowcases [40].

Beyond apparel and home textiles, cotton is essential in industrial uses. It is employed in manufacturing medical supplies such as bandages, surgical gowns, and swabs due to its natural absorbency and softness. Cotton fibers are also used to produce industrial threads, tarpaulins, zipper tapes, bookbindings, and wall coverings. Additionally, cotton is integral to technical textiles, where it is treated to become flame-resistant or chemical-resistant for protective clothing used in industries like firefighting, welding, and construction. Cotton fabrics are also used in filtration systems because of their high absorbency and breathability [36].

The fiber's natural properties combined with ongoing innovations, such as cationic treatments that improve dye uptake and reduce environmental impact, ensure cotton remains a preferred material in both consumer and industrial textile applications.

Its socio-economic importance and adaptability continue to make cotton a cornerstone of the global textile industry, balancing comfort, durability, and sustainability [39]. *Gossypium hirsutum* (upland cotton) production varies globally and is influenced by cultivar selection, environmental conditions, and management practices.

Studies have shown significant differences in growth and yield among various cultivars, with traits such as plant height, number of sympodial branches, boll number, boll weight, seed cotton yield, and fiber quality parameters differing notably (e.g., FH-115, FH-207, MNH-786).

For instance, cultivars like Edessa and PG 2018 have demonstrated superior seed cotton yield and fiber yield, maintaining stability across different years and environmental conditions [41].

Cotton yield potential can reach theoretical maximums of around 5000 kg lint per hectare under optimal long-season growing conditions. Management systems such as organic, biodynamic, Bt-conventional, and non-Bt conventional cultivation also impact growth and yield attributes, with Bt-conventional systems often showing higher seed cotton yields.

In semi-arid regions, varieties like RH-668 have been developed for heat and drought tolerance, showing improved yield and fiber quality compared to standard varieties, making them suitable for challenging climates. Sustainable cotton production also requires balanced nutrient management, including nitrogen, phosphorus, and potassium, to maintain soil fertility and achieve high yields.

Overall, the production of *Gossypium hirsutum* depends on selecting high-yielding, climate-resilient cultivars combined with appropriate agronomic practices to optimize yield and fiber quality in diverse ecological zones[42].

2.2 Factors Contributing to Low Cotton Yield in the Country

2.2.1 Cultivator Selection

One of the most crucial factors influencing cotton yield is the selection of an appropriate cultivar that is well-suited to the specific agro-climatic conditions of a region. The success of cotton production heavily depends on using varieties developed and recommended by scientists for their adaptability to environmental factors such as temperature, light, wind, and humidity. Certified seeds, whether from public or private sources, undergo field inspections and laboratory testing by seed certification authorities to ensure they meet standards for varietal purity and are free from weed seeds, other crop admixtures, and diseases [43].

Despite these standards, many farmers continue to use non-recommended and uncertified seeds. These often exhibit poor germination rates, and even when germination is adequate, the plants tend to produce fewer buds, flowers, and bolls at maturity. In some cases, sowing such seeds results in fields with a mix of plant types some tall, others dwarf with inconsistent characteristics. Additionally, substandard seeds often come with heavy insect infestation, further reducing crop quality and yield potential [44].

2.2.2 Soil Preparation

Cotton thrives best in soils that offer excellent water retention and aeration, combined with efficient drainage, as the crop is sensitive to waterlogging and excessive moisture. For optimal plant growth and development, the soil must maintain adequate moisture levels and temperature while minimizing resistance to root penetration. This can be achieved through deep ploughing, which loosens the soil and supports healthy root development [45].

An effective tillage system plays a critical role in preparing an ideal seedbed, providing suitable conditions such as proper temperature, moisture, and minimal soil compaction for seed germination and early plant growth. Proper land leveling further enhances efficiency by ensuring uniform water distribution during irrigation and reducing the overall use of agricultural inputs [42].

Various studies and practical experiences have shown that deep tillage reduces subsoil compaction and helps conserve soil moisture. Among different planting methods, bed planting has shown better results compared to ridge and flat planting systems. Ridge tillage, in particular, has been noted to improve both soil physical properties and crop performance.

Different tillage practices can significantly influence cotton yield, with deep ploughing consistently producing higher yields compared to minimal tillage. Overall, effective soil preparation ensures unimpeded root growth and establishes the ideal environment for healthy plant development and maximum productivity [43].

2.2.3 Seed Rate and Plant Spacing

Using the correct seed rate is essential for achieving healthy plant growth and maximum yield. The ideal seed rate varies depending on factors such as cotton variety, soil condition, sowing technique, and overall farming practices. For genetically pure seeds with high germination potential, a seed rate of 15-25 kg per hectare is generally recommended based on extensive research and field experience. This rate can be adjusted within various cropping systems without negatively affecting yield or complicating crop management [46].

However, many farmers continue to use a lower-than-recommended seed rate, which leads to insufficient plant populations in the field and ultimately results in reduced yield. Alongside seed rate, proper plant spacing also plays a vital role in cotton production.

Both overcrowding and sparse planting can negatively impact yield per hectare.The recommended spacing for cotton is 30 cm between plants, with 75 cm between

rows, depending on whether the selected variety has a bushy or compact growth habit. Maintaining appropriate spacing ensures optimal plant development and higher productivity. In some cases, bed planting with specific plant spacing, such as 22.5 cm, has been found effective in achieving optimal plant populations and improving yield [47].

2.2.4 Sowing/Planting Date

The timing of cotton sowing plays a vital role in determining crop performance and yield. Delayed sowing is one of the primary causes of reduced yield, while sowing too early often leads to poor germination and weak crop establishment. On the other hand, planting too late results in excessive vegetative growth, making the crop harder to manage and ultimately reducing yield [48].

For successful cotton cultivation, adhering to the recommended sowing window is crucial, as it helps minimize the impact of adverse environmental factors. However, delays in sowing are commonly observed at the farmer level due to various challenges, such as the unavailability of certified seed, water shortages, and lack of timely access to fertilizers. These issues contribute to poor crop growth and significant yield losses.

To ensure optimal growth, boll formation, and fiber quality, it is important to follow the proper sowing schedule. Research and field experience have shown that planting cotton within the ideal timeframe enhances plant development, boll number, boll weight, and overall yield. Sowing outside the optimal window either too early or too late—has consistently been associated with a steady decline in yield and deterioration in fiber characteristics such as staple length, strength, and maturity. Therefore, timely planting is a key agronomic practice for achieving high productivity and quality in cotton production [49].

2.2.5 Irrigation

Irrigation plays a role just as vital as fertilizers and tillage in supporting crop production. It provides essential moisture to the crop, especially in regions where rainfall is insufficient. A lack of water or prolonged dry spells during the growing season can cause substantial yield losses and negatively impact farmers' income. To maintain sustainable productivity, it is important to supply water consistently and according to the crop's specific needs throughout its lifecycle [50].

Missing irrigation during one or more critical growth stages such as flowering or boll formation can lead to a significant drop in overall yield. Studies and practical farming experience have demonstrated that proper irrigation not only increases yield but also improves crop quality and profitability. In most cases, around six well-timed irrigations are found to be ideal for achieving maximum cotton yield [51].

The amount and timing of water application greatly influence boll development and seed cotton production. Full irrigation at all key growth stages typically results in higher yields, while water shortages or irregular irrigation schedules can severely affect crop performance. Therefore, adopting a well-planned irrigation strategy is essential for optimizing cotton yield and ensuring the crop reaches its full potential [52].

2.2.6 Nutrients

Most agricultural soils, particularly in cotton-growing regions, tend to have low levels of organic matter. In addition, nutrient deficiencies are widespread due to repeated cultivation of nutrient-exhaustive crops, high temperatures, low rainfall, high fertilizer costs, and imbalanced fertilizer use. These factors collectively contribute to poor soil fertility and reduced crop productivity [53].

Applying fertilizers in the right proportions, using appropriate methods, and timing their application according to soil nutrient status, moisture levels, crop type, and growth stage can significantly improve yield by as much as 25% to 75%.

Among the essential nutrients, nitrogen, phosphorus, and sometimes potassium are often deficient, leading to poor plant growth and lower yields. Nitrogen plays a key role in plant development as it is a crucial component of chlorophyll, nucleic acids, proteins, and growth regulators [54].

Proper nitrogen management is especially important in cotton cultivation, as both deficiency and excess can adversely affect yield. While nitrogen improves plant vigor and boll development, imbalanced application may lead to excessive vegetative growth at the expense of reproductive development. Phosphorus and potassium also play vital roles in crop growth and yield formation. However, overuse of these macronutrients can disrupt the uptake of important micronutrients, leading to secondary nutrient deficiencies.

In many areas, cotton shows a strong and consistent response to nitrogen application across all soil types, while phosphorus responses tend to vary based on soil conditions. Balanced fertilizer use, especially with an appropriate combination of nitrogen, phosphorus, and potassium (N: P: K), has been shown to maximize seed cotton yield. Moreover, micronutrients such as zinc and boron, when applied as foliar sprays, further enhance cotton yield and improve crop performance [55].

2.2.7 Crop Protection

One of the major challenges throughout the cotton growing season is the threat posed by weeds, insect pests, and diseases. These issues can cause significant economic losses by lowering both crop yield and fiber quality. Managing pests often requires substantial investment in pesticides and other weed control measures, which becomes a financial burden for cotton growers. However, several factors reduce the effectiveness of these protective measures, including the use of low-quality or adulterated products, high costs, lack of availability at critical times, and the use of faulty equipment by untrained laborers. Moreover, the limited awareness among farmers about proper pest control practices further contributes to crop loss [56].

2.2.8 Weeds

Weeds compete aggressively with cotton plants for essential resources such as light, water, and nutrients, especially during the early growth stages. If not controlled soon after emergence, this competition can lead to a significant reduction in yield. In addition to direct competition, some weeds also act as hosts for harmful pests, diseases, and nematodes, increasing the risk of crop damage.

Weeds have a particularly high demand for key nutrients like potassium, nitrogen, and magnesium often absorbing several times more than the crop itself. This nutrient drain, along with their interference with light and water absorption, makes them a serious threat to cotton productivity.

Effective weed control, especially through timely herbicide application and interculturing practices, has been shown to enhance seed cotton yield. Pre-emergence herbicide treatments are particularly effective in reducing weed density and promoting better crop growth. Conversely, poor weed management leads to dense infestations, which drastically lower cotton yield [57].

2.2.9 Insects

Insect pests are one of the primary threats to cotton production, causing serious reductions in yield and quality. They can damage both the vegetative and reproductive parts of the plant. When the vegetative parts are affected, plant growth is delayed or weakened, while attacks on the reproductive structures directly reduce the number of bolls, resulting in lower harvests. Some pests also defoliate the plant, which decreases boll size, weakens the plant, and diminishes fiber quality [58].

Among the most damaging pests are sucking insects such as jassids, whiteflies, and thrips, which are responsible for significant yield losses. If not controlled effectively, these pests can drastically reduce the number of bolls formed and the overall health of the plant.

Bollworms also pose a serious risk, often leading to substantial losses during key reproductive stages. Efficient pest management strategies, including timely application of insecticides and integrated pest management practices, are essential to prevent severe crop damage and maximize productivity [59].

2.2.10 Diseases

Diseases caused by fungi, bacteria, and viruses are another major factor contributing to reduced cotton yield. These pathogens can affect cotton at various stages, causing issues such as boll rot, root rot, and cotton leaf curl virus (CLCuV). Infected plants often show stunted growth, poor boll formation, and diminished fiber quality. Harvesting diseased plants along with healthy ones further reduces the overall quality of the cotton [60].

Effective disease management begins with the development and use of resistant or tolerant cultivars. Other important measures include maintaining good field hygiene, proper irrigation scheduling, optimal planting time, and avoiding excessive use of organic manures that may promote excessive vegetative growth. Cultural practices like compost incorporation improve soil health and help beneficial microorganisms suppress disease agents. These microbes may compete for nutrients, produce natural antibiotics, prey on pathogens, or even trigger the plant's own disease resistance mechanisms.

Chemical treatments with fungicides and biological control methods can also play a role in managing disease pressure. Some diseases, like fusarium wilt and boll rot, can cause severe losses, especially in humid conditions. Cotton leaf curl virus (CLCuV), spread by whiteflies, has emerged as one of the most destructive viral diseases, leading to major yield reductions by severely affecting plant height, boll formation, and overall growth [61].

2.2.11 Modern Technology

The adoption of modern agricultural technology is crucial for improving cotton yield and transitioning from traditional subsistence farming to a more market-oriented approach. New technologies enhance efficiency at every stage, from cultivation and crop protection to harvesting and marketing. Innovations in plant protection, including advanced pest and disease control techniques, can significantly reduce crop losses and improve productivity.

Farmer education and skill development are essential components of successful technology adoption. Well-informed farmers are better equipped to implement modern practices, select high-performing cultivars, manage resources efficiently, and stay updated on market trends. The ability to understand and use modern tools, from improved irrigation systems to advanced fertilizers and mechanized harvesting equipment, empowers farmers to boost production while reducing costs and losses [62].

2.3 Genomic Features

Due to hybridization between two diploid progenitor species, *G. arboreum* (A-genome donor) and *G. raimondii* (D-genome donor), the genome of *G. hirsutum* is allopolyploid (2n = 4x = 52). The coexistence of two subgenomes within the same nucleus as a result of this allopolyploid origin has complicated gene expression, regulation, and evolution.

Thanks to developments in next-generation sequencing, the genome of *G. hirsutum* has been assembled in high resolution, revealing its size (∼2.5 Gb), gene content (∼70,000 genes), and structural characteristics like segmental duplications, transposable elements, and extensive synteny with related species. Numerous gene families implicated in important agronomic traits, such as those controlling fiber development, stress responses, and flowering time, have been found through genome mapping efforts [63].

The *SAP* gene family is one of the MADS-box transcription factors that play a crucial regulatory role in floral organ specification within this genomic context. Through comparative analyses with model species, studies have mapped these genes, annotated their structures, and predicted their functions by utilizing genomic resources. When polyploidy produces homeologous gene pairs, it is important to distinguish between duplicate copies and their possible sub functionalization or neo functionalization. Comprehensive analyses of *G. hirsutum* gene families have been made easier by bioinformatics tools [64].

2.4 Floral Morphology and Development

Floral development in *G. hirsutum* is integral to its reproductive success and yield. The typical cotton flower is composed of five petals forming a conspicuous corolla, five stamens fused into a staminal column, and a central pistil. This arrangement is characteristic of *Malvaceae*, reflecting evolutionary conservatism in floral architecture.

The ABCDE model of flower development, which postulates combinatorial interactions among homeotic genes specifying organ identities across concentric whorls, controls the development of these floral organs. Sepals are specified by A-class genes, petals by A+B, stamens by B+C, carpels by C alone, ovule identity by D, and these functions are supported by E-class genes as crucial cofactors.

SAP genes are thought to be B-class interactors or belong to the B-class MADS-box transcription factors, and they are important for determining the identity of the stamen and petals. Reproductive fitness may be impacted by homeotic changes brought on by mutations or changed expression of *SAP* genes, such as the loss of petals or the transformation of stamens into carpeloid structures [64].

The importance of floral morphology extends beyond reproductive assurance. In cotton breeding, flower structure affects boll formation, pollen viability, and susceptibility to environmental stresses such as high temperatures that disrupt pollen development.

Breeding for optimal floral traits is thus a critical component of yield improvement programs. Studies of floral development in *G. hirsutum* have employed both classical genetic approaches and modern molecular tools. Breeding for optimal floral traits is thus a critical component of yield improvement programs. Studies of floral development in *G. hirsutum* have employed both classical genetic approaches and modern molecular tools.Breeding for optimal floral traits is thus a critical component of yield improvement programs. Studies of floral development in *G. hirsutum* have employed both classical genetic approaches and modern molecular tools.

TTranscriptome analyses have provided expression profiles of key developmental regulators across flower stages, revealing dynamic gene networks that respond to environmental cues. Such studies underscore the value of integrating molecular genetics, genomics, and bioinformatics to unravel the complexities of cotton floral biology ranscriptome analyses have provided expression profiles of key developmental regulators across flower stages, revealing dynamic gene networks that respond to environmental cues. Such studies underscore the value of integrating molecular genetics, genomics, and bioinformatics to unravel the complexities of cotton floral biology [65].

2.5 Mechanism of Cotton Resistance to Abiotic Stress

Abiotic stress refers to environmental challenges such as extreme temperatures, drought, salinity, alkalinity, and low temperatures that negatively affect plant health and development. These stresses trigger a range of physiological and biochemical reactions in plants as they attempt to adapt and survive under harsh conditions. In cotton, as in many other crops, these adverse environmental factors impact not only growth and development but also the plant's morphology, physiology, and overall productivity, including fiber quality and yield.

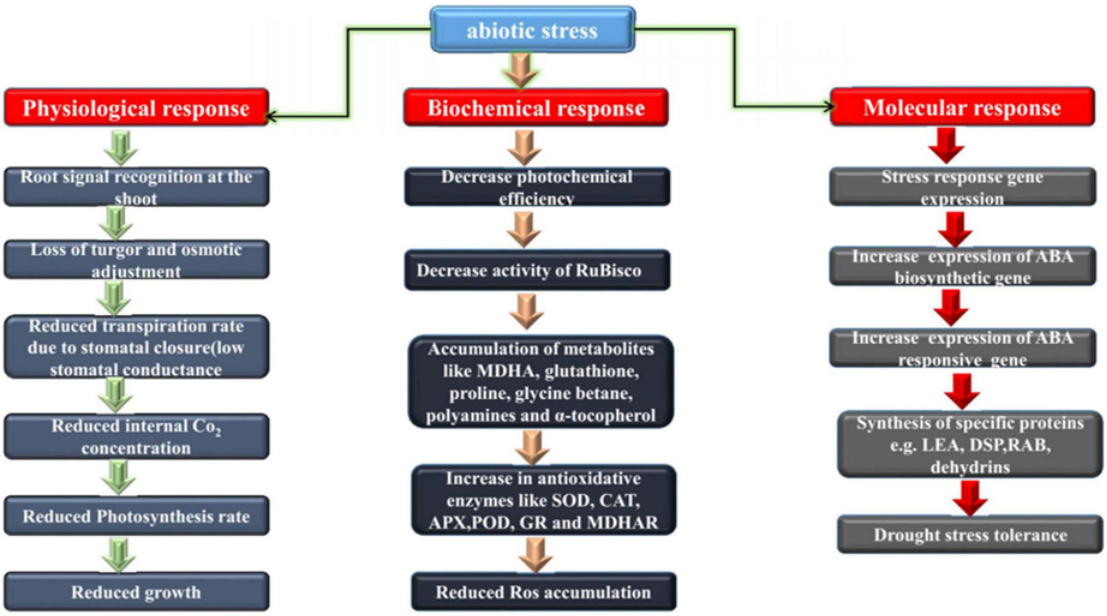

FIGURE 2.5: Physiological, biochemical and molecular basis of drought resistance in plants [66].

Understanding how plants respond to abiotic stress has been a major focus of research. Initial studies focused on model plants, and over time this research has expanded to include important crops such as rice, maize, soybean, and cotton. Through long-term evolution, plants have developed various strategies to resist or tolerate environmental stresses. These resistance mechanisms are not governed by a single process, but rather involve complex, multi-layered systems that include molecular, physiological, and biochemical components [66].

Advances in molecular biology have significantly deepened our understanding of plant responses to stress. These developments have revealed that stress resistance involves changes at the genetic level, including gene expression regulation and signal transduction pathways. In cotton, abiotic stress responses are now understood to involve four main phases:

1. *Perception Phase* - the plant detects the presence of environmental stress.
2. *Signal Transduction Phase* - once the stress signal is detected, it is transmitted throughout the plant's internal signaling network.
3. *Response Phase* - the transmitted signal triggers a cascade of molecular responses.

4. *Gene Expression Phase* - resistance-related genes are activated to initiate defense mechanisms.

Once these stages are completed, cotton plants demonstrate specific physiological and molecular responses that help them resist the effects of abiotic stress [67].

Given the serious impact of environmental stress on cotton production, it is essential to study and enhance cotton's ability to adapt. Developing stress-tolerant cotton varieties through genetic improvement and breeding is not only scientifically significant but also critical for sustainable cotton cultivation. Ongoing research focuses on exploring resistant germplasm resources, understanding their resistance mechanisms, and applying this knowledge to future breeding programs aimed at improving cotton's resilience in challenging environments [68].

2.5.1 Effects of Non-Living Stress on Cotton Growth and Development

Non-living, or abiotic, stresses refer to harmful environmental conditions such as drought, salinity, extreme temperatures, nutrient-poor soils, and other unfavorable factors. These conditions can severely disrupt the normal physiological and developmental processes of cotton plants. When cotton is exposed to intense abiotic stress, its growth is often stunted, and the ability to develop properly is compromised. In extreme cases, these stresses can lead to a significant drop in yield or even result in the death of the plant [69].

2.5.2 Effects of Abiotic Stress on Cotton Growth and Development

Among all non-biological stress factors affecting cotton, drought is considered the most harmful and ranks as one of the leading causes of crop loss. Drought stress leads to severe dehydration of plant cells, disrupting the normal structure of cell membranes. This damage often causes the stomata tiny openings on leaf surfaces

to close excessively, limiting the intake of carbon dioxide and reducing the rate of photosynthesis [70].

Drought also negatively impacts seed germination in cotton. Under water-deficient conditions, reductions are observed in water potential, germination rate, seedling height, root length, root-to-stem ratio, as well as the dry and fresh weight of young seedlings. These effects hinder early growth and establish weak plant development.

In field studies, drought stress has been shown to slow cotton growth significantly, particularly in the early stages, with noticeable declines in plant height and leaf emergence. The severity of the impact varies depending on the growth stage at which the plant experiences drought. The most critical stages affected are the flowering and boll formation stages, followed by the bud, maturity, and seedling stages. However, some drought-resistant cotton varieties have shown the ability to adapt better under these conditions. These plants can minimize water loss by increasing resistance in their stomata and lowering the rate of water evaporation, which helps maintain efficient photosynthesis. Additionally, drought-tolerant cotton types often exhibit increased root activity and higher accumulation of proline an important compound that aids in stress resistance enhancing their ability to cope with water scarcity [71].

2.5.3 Effects of Salt and Alkali on Cotton Growth and Development

Salt damage is one of the major challenges in agriculture, particularly affecting crops like cotton. While low levels of salt in the soil can actually serve as nutrients and support healthy plant growth, excessive salt concentrations become harmful. When salt levels are too high, they can cause premature aging or decay of leaves, ultimately leading to reduced cotton yield. Salt stress typically becomes problematic when soil salinity exceeds a certain threshold. At that point, it triggers a series of harmful effects on cotton, including ion toxicity, osmotic stress, and oxidative stress. These stresses can impair plant functions, hinder development, and in severe cases, result in plant death. High salt levels in the soil can interfere

with water absorption, causing swelling of seeds and slowing down germination. As a result, seed germination efficiency drops, and the young plants become weak. In the roots, salt-induced infiltration stress can lead to cellular dehydration and loss of turgor pressure, disrupting essential physiological processes [72].

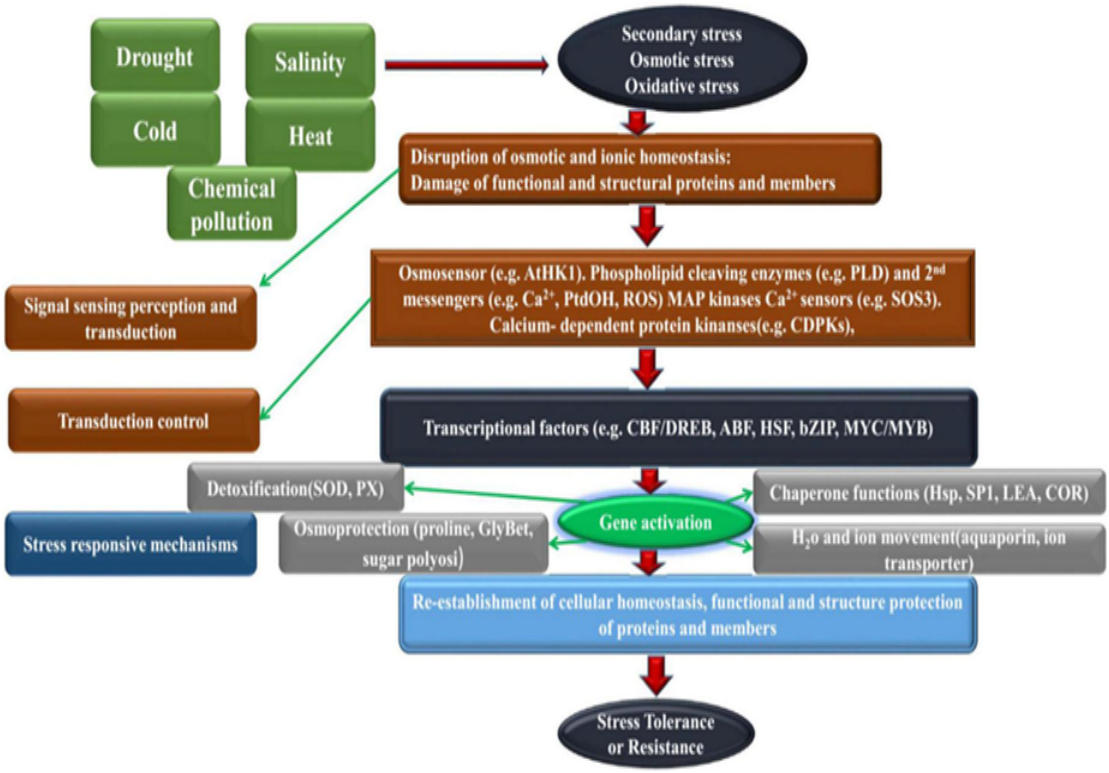

FIGURE 2.6: The complexity of plant response to adversity stress [72].

The severity of osmotic stress increases with higher salt concentrations. Additionally, excessive salt ions create competition among nutrients, limiting the plant's ability to absorb essential minerals. For example, salt stress especially from sodium chloride can significantly reduce levels of key nutrients like calcium, potassium, magnesium, phosphorus, and manganese in both leaves and roots, leading to nutritional imbalances [72].

Over time, cotton plants have developed mechanisms to cope with saline environments. These adaptive strategies generally involve regulating salt ion concentrations and managing osmotic balance. A high internal ratio of potassium to sodium (K^+/Na^+) and calcium to sodium (Ca^{2+}/Na^+) is often an indicator of better salt tolerance. Salt-tolerant cotton varieties tend to maintain higher K^+/Na^+ ratios compared to salt-sensitive ones. In salt-tolerant plants, potassium and calcium

ions are more effectively retained in the leaves, whereas sensitive plants accumulate more chloride ions. This difference is largely due to selective ion transport, especially the root's ability to selectively absorb potassium over sodium, helping maintain better nutrient balance under stress [73].

Betaine, a naturally occurring compound in plants, also plays a key role in helping cotton survive saline conditions. It accumulates in plant cells during stress, helping to lower osmotic pressure and protect important cell structures. Furthermore, cotton plants have a highly active system for removing lipid peroxides, which helps maintain the integrity of cell membranes and supports energy balance. In salt-tolerant cotton varieties, this protective system is significantly more active. They show increased activity of enzymes like glutathione reductase and higher levels of protective compounds, which are essential for maintaining normal cellular functions. In contrast, salt-sensitive varieties do not show such significant changes, leaving them more vulnerable to damage under saline conditions [74].

2.5.4 Influence of Temperature Change on Cotton Growth

Temperature fluctuations, whether extremely low or high, can significantly affect the growth, development, and overall productivity of cotton. Cotton is particularly sensitive to both cold and heat stress, and each type of temperature stress impacts the plant in unique ways.

Low temperatures, including cold and frost damage, can slow down or completely halt cotton's metabolic activity. When exposed to cold stress, cotton plants may suffer partial or complete wilting, depending on the severity and timing of the temperature drop. The seedling stage is especially vulnerable if frost damage affects more than 15% of the plant, it can delay the entire growing cycle, ultimately reducing both yield and fiber quality. Under cold conditions, seed germination slows drastically, and in some cases, seeds may even rot. Even if germination occurs, the growth of shoots is stunted, resulting in weak and fragile seedlings. Additionally, low temperatures combined with limited sunlight interfere with the plant's ability to perform photosynthesis. This happens because the energy transfer between the

photosystems involved in capturing light energy is disrupted. One way cotton can show some tolerance to cold is through increased nitrate reductase activity, which helps boost nitrogen content in plant tissues, making them more resilient [75].

Although cotton originated in tropical regions, its ability to tolerate high temperatures has not fully developed despite years of domestication. Heat stress can negatively affect seed germination and severely inhibit root development. While cotton seeds germinate best between 28–30^{o}C and roots grow well between 22–35^{o}C, temperatures above 40^{o}C during the day and 32^{o}C at night are considered stressful. Such heat levels can reduce germination rates, hinder water and nutrient absorption, and impair the development of both tap and lateral roots even when moisture and nutrients are available. As a result, the root system becomes weak and shallow [76].

High temperatures also speed up the plant's lifecycle, causing it to mature more quickly than normal. This reduces the time available for boll development and carbon assimilation, leading to smaller bolls, decreased yields, and lower-quality fiber. Moreover, elevated night temperatures can affect the formation and quality of buds and seed pods. During the reproductive phase, cotton poll High temperatures also speed up the plant's lifecycle, causing it to mature more quickly than normal. This reduces the time available for boll development and carbon assimilation, leading to smaller bolls, decreased yields, and lower-quality fiber. Moreover, elevated night temperatures can affect the formation and quality of buds and seed pods. During the reproductive phase, cotton pollen is extremely sensitive to heat. Pollen viability, germination rate, and pollen quantity decline significantly when temperatures rise, resulting in poor fertilization and pollen abortion. There is an optimal temperature for pollen tube growth, and when this limit is exceeded, reproductive success is compromised [77].

In response to high heat, cotton like other plants produces special proteins known as heat shock proteins (HSPs). These molecules help protect cells from damage and support recovery. Enhancing the production of these proteins is one strategy plants use to strengthen their resistance to heat stress, helping them maintain cellular function and survive in harsh environments.

2.5.5 Effects of Soil Nutrient Stress on Cotton Growth and Development

Soil nutrients like nitrogen, phosphorus, and potassium are fundamental for plant life, and they play a crucial role in determining cotton yield and fiber quality. When these nutrients are deficient, the growth and development of cotton plants are significantly impaired [76].

Potassium, for instance, is essential for root development and overall plant vigor. A shortage of potassium hampers the elongation of roots and restricts lateral root formation, limiting the plant's ability to absorb water and nutrients. It also affects stem thickening, vegetative growth, and reduces the chlorophyll content and photosynthesis rate in leaves.

Prolonged potassium deficiency can lead to early signs of physiological decline such as red petiole rot, reduced reproductive growth, and increased vulnerability to diseases. On the other hand, cotton varieties that are efficient in potassium use tend to perform better under potassium-deficient conditions. These varieties transport potassium more effectively to above-ground parts, ensuring healthy chloroplast function in leaves. They also develop more root hairs near the root tip, which enhances their nutrient uptake. Improvements in xylem differentiation and stronger root activity help maintain membrane stability and overall resistance under stress [77].

Phosphorus is another critical nutrient, and its deficiency affects cotton at multiple levels. Under phosphorus stress, plants show stunted growth, smaller leaves, weaker stems, poor root development, delayed flowering, and low boll formation. This results in reduced yield and inferior fiber quality.

In response, cotton plants adapt by modifying root architecture lateral roots and root hairs become more prominent, increasing the root surface area available for phosphorus uptake. The plasma membrane of root cells plays a central role in this process by facilitating phosphorus transport into the plant.

Specialized transporter proteins located in different cellular membranes help move phosphorus from the soil into the cells and then to various organelles where it is needed.Nitrogen, too, is vital for plant metabolism and growth. Although large quantities of nitrogen fertilizer are used in agriculture, much of it is lost due to inefficient uptake and usage.

Cotton plants vary in their ability to absorb and utilize nitrogen efficiently. Under low nitrogen conditions, more efficient varieties are able to adjust by increasing the active absorption area of roots. They develop more extensive root systems, which improves their ability to take up available nitrogen from the soil. Additionally, they release protons and organic acids into the rhizosphere, which enhances nutrient availability.

These changes help in better nitrogen utilization even when external supplies are limited. When nitrogen is available, the plant directs most of it toward the leaves, followed by stems and roots. This distribution supports photosynthesis and overall plant productivity [76].

In modern cotton breeding, there is a growing focus not just on enhancing yield and disease resistance but also on improving tolerance to abiotic stress and increasing nutrient use efficiency.

Developing nutrient-efficient cotton varieties ensures sustainable production even in nutrient-poor soils, making it a vital goal for future agricultural success [79].

2.6 Transcription Factors and Plant Growth

Transcription factors (TFs) play a vital role in guiding plant growth, development, and adaptation by responding to various internal and external signals. They act as master regulators, controlling the expression of key genes involved in major developmental pathways. Throughout a plant's life cycle, from germination to maturity, TFs ensure that specific genes are turned on or off at the right time [80].

One of the earliest and most critical stages in a plant's life is seed germination, which is tightly regulated at the molecular level. Alongside plant hormones and dormancy-related mechanisms, transcription factors are actively involved in orchestrating this process. Several TF families contribute to seed germination, including bZIP, MADS-box, NF-YC, and DOF. These factors regulate gene expression patterns necessary for initiating and maintaining germination [81].

Additionally, TFs like *FUSCA3* (a B3-domain protein) and *GATA12* (a zinc finger TF) are important for maintaining seed dormancy, especially in model plants like *Arabidopsis thaliana*. Other transcription factors such as TCP14 and TCP15, which belong to the bHLH family, help regulate seed germination by influencing gibberellin-related pathways. They do this by interacting with promoters of specific genes like EXPANSIN, thereby affecting hormone signaling and cell wall remodeling crucial for seedling emergence [82].

TABLE 2.1: Transcription factors identified for their role in plant development [83].

Plant System	**Transcription Factors**	**Function**
Arabidopsis thaliana	AtWRKY75	Overexpression in Arabidopsis thaliana speeds up the onset of flowering.
A. thaliana	AtTCP	Expression in Arabidopsis thaliana enhances leaf development.
A. thaliana	AtAP2/ERF2 (DRNL)	Expression of AtAP2/ERF2 (DRNL) in Arabidopsis thaliana promotes gynoecium development.
A. thaliana	AtGIS	Expression in Nicotiana tabacum controls the development of glandular trichomes.
A. thaliana	AtWOX11/12	Expression in Arabidopsis thaliana triggers root primordia formation and root organogenesis.
A. thaliana	AtLBD29	Expression in Arabidopsis thaliana promotes lateral root emergence.
Brassica rapa	BrZFP38	Expression in Brassica rapa promotes flower development by regulating key floral genes and pathways.
Glycine max L.	GmNAC004	Overexpression in Arabidopsis thaliana enhances lateral root development.

continued on next page

Table 2.1 continued from previous page

Plant System	Transcription Factors	Function
Oryza sativa	OsTCL1	Overexpression in Arabidopsis thaliana enhances lateral root development.
Solanum lycopersicum L.	SlWUS	Expression in Solanum lycopersicum promotes flower development.

Transcription factors (TFs) also play a crucial role in enhancing seed germination, especially under stress conditions. For example, certain TFs help seeds recover and resume growth after exposure to high salinity. One such factor is OsbHLH035, which supports germination and seedling recovery by regulating pathways linked to salt stress tolerance. This regulation involves both ABA-dependent and independent mechanisms, helping the plant adapt more effectively to challenging environments [84].

2.6.1 Floral Development

Beyond germination, TFs are central to the reproductive development of flowering plants. Flowering, a key event in a plant's life cycle, is tightly controlled by various TF families, including MYB, MADS, and WRKY. These factors ensure the precise timing and coordination required for successful reproduction. In some plants like chrysanthemums, specific MYB-type TFs regulate flowering onset. In *Arabidopsis*, MYB30 promotes flowering by interacting with core flowering genes, ensuring the plant transitions efficiently into the reproductive phase under different day lengths. Similarly, in wheat, the TF TaMYB72 has been linked to earlier flowering, even when introduced into rice [85].

Interestingly, not all MYB TFs promote flowering. Some, such as MYB44 and PtrMYB192, actually delay it, showing that members of the same TF family can have opposite effects depending on the context. MYB proteins also work in combination with other TFs. For instance, in alfalfa, SPL13 affects flowering time by suppressing a MYB gene involved in the process. Apart from timing, MYB

TFs are also involved in determining flower color. In tree peony, a particular MYB factor promotes the production of anthocyanins, the pigments responsible for vivid flower colors [86].

MADS-box transcription factors are another important group involved in floral development. These TFs regulate key changes in the shoot apical meristem as it shifts from vegetative growth to flowering. For instance, some MADS-box genes promote flowering regardless of the photoperiod, indicating their powerful role in developmental transitions [87].

In bamboo, a MADS-box gene is responsible for switching the plant from growth to reproductive mode. Other important TFs like *APETALA1* (*AP1*) and *LEAFY* (*LFY*) are essential for the formation of floral meristems, ensuring proper flower structure and function. Together, these transcription factors form a complex and coordinated network that drives the critical stages of plant life [88].

2.6.2 Leaf Development

Leaves are the main sites of photosynthesis in plants and play a vital role in determining overall plant productivity. The development of leaves is influenced by various factors including the species of the plant, its growth stage, and surrounding environmental conditions. However, at the molecular level, leaf development is primarily controlled by a complex interplay of plant hormones, transcriptional regulators, and the mechanical properties of plant tissues [8].

Among the key transcription factors involved in leaf development are the GRF (GROWTH-REGULATING FACTORS), TCP, and ZIP families. GRFs, in particular, promote the growth of lateral organs and are essential for normal leaf formation. Other families of transcription factors also contribute significantly.

For instance, the *WOX* (*WUSCHEL-RELATED HOMEOBOX*) and YABBY families are responsible for establishing the upper (adaxial) and lower (abaxial) sides of the leaf. This polarity is crucial for proper leaf structure and function [90].

CIN-TCP transcription factors are also instrumental in shaping leaf patterns and determining their final form. Additionally, the *CUP-SHAPED COTYLEDON* (*CUC*) family of TFs plays a role in forming and organizing the boundaries between plant organs, further supporting leaf development [91].

2.6.3 Root Development

Roots are essential for absorbing water and nutrients from the soil and anchoring the plant securely. The development and growth of roots are regulated by several transcription factor families, including NAC, MYB, bHLH, and MADS-box [92].

For instance, NAC1 is important for initiating lateral root formation, a process critical for increasing the root surface area and improving nutrient uptake. MADS-box TFs, along with other factors like ANR1, help control root growth in response to nutrient availability, especially nitrates in the soil [93].

Root hair formation—tiny projections from root cells that further enhance nutrient and water absorption—is also regulated by a network of transcription factors.

These include members of the bHLH family such as *GLABRA3* (*GL3*) and *ENHANCER OF GLABRA3* (*EGL3*), as well as *GLABRA2* (*GL2*) from the HD-ZIP family. TFs from the MYB family, like CAPRICE (CPC), TRY, WER, and ETC1, also play a pivotal role in determining where and how root hairs develop [94].

2.6.4 Transcription Factors and Stress Responses in Plants

Plants are constantly exposed to a range of environmental stresses, both biotic (such as pathogens and pests) and abiotic (such as drought, heat, salinity, and cold). To survive and adapt under these conditions, plants rely on transcription factors (TFs) that activate stress-responsive genes. These TFs bind to specific promoter regions of these genes and modulate their expression, helping plants initiate protective and adaptive responses [95].

2.6.5 Major TF Families Involved in Stress Responses

Among the many transcription factor families in plants, several have been consistently recognized for their crucial roles in managing stress responses. These include:

a) AP2/ERF (APETALA2/Ethylene Responsive Factor)

b) bHLH (Basic Helix-Loop-Helix)

c) MYB (Myeloblastosis-related)

d) NAC (No Apical Meristem, ATAF1/2, and CUP-SHAPED COTYLEDON2)

e) WRKY

f) bZIP (Basic Leucine Zipper)

g) HSFs (Heat Shock Factors)

Each of these families has members that are actively involved in regulating plant defense and adaptation to stress conditions [96].

2.6.6 Role of NAC Transcription Factors

NAC TFs are particularly known for their involvement in drought and salt stress responses in multiple plant species. Some NAC members help regulate tolerance to multiple stresses, including heat and oxidative damage, by modulating hormone pathways and maintaining reactive oxygen species (ROS) balance [97]. Overexpression of certain NAC TFs in model plants has led to enhanced resistance against harsh environmental conditions, such as drought, salinity, and heat [98].

2.6.7 MYB Transcription Factors and Stress Regulation

MYB transcription factors are key players in both abiotic and biotic stress responses. They regulate a variety of physiological processes, including drought resistance, by controlling genes related to water retention and wax biosynthesis [99]. In some cases, their expression is strongly induced under stress treatments such as salt and drought, further confirming their regulatory role. MYB TFs are also associated with managing temperature stress across different plant species [100].

2.6.8 WRKY TFs in Stress Defense

WRKY TFs are widely known for their involvement in plant immune responses and adaptation to environmental challenges. They contribute to plant resistance under salt, heat, and drought stress by activating relevant stress-related genes. These TFs are active across a range of species, playing roles in both localized and systemic stress responses [101].

2.6.9 Heat Shock Factors (HSFs)

HSFs primarily regulate temperature-related stress responses. These TFs become active during extreme heat conditions and support the expression of heat shock proteins (HSPs), which protect plant cells from thermal damage. Their function is well-documented in several plant species, where they enhance thermo tolerance and ensure survival under fluctuating temperatures [102].

2.6.10 *AP2/ERF* Family in Temperature Stress

Members of the *AP2/ERF* family are also closely associated with temperature stress tolerance. These TFs are involved in regulating cold and heat responses and help coordinate hormonal and metabolic pathways that support plant adaptation under temperature extremes [103].

TABLE 2.2: Exemplification of the contribution of transcription factors in stress challenged plants [104].

Plant System	**Transcriptions Factors**	**Stress**	**Function**
Arabidopsis thaliana	AtWRKY30	Heat / Drought	Overexpression in *Triticum aestivum L.* improves stress tolerance by enhancing the antioxidant system, increasing chlorophyll content, & upregulating stress-responsive genes, while simultaneously reducing electrolyte leakage, hydrogen peroxide, & malondialdehyde content.
A. thaliana	AtWRKY11 & AtWRKY17	ABA/salt/	Expression in *Arabidopsis thaliana* improves stress tolerance by increasing germination rates & promoting root development.
A. thaliana	AtJUB1	Drought	Overexpression in Solanum lycopersicum L. enhances stress tolerance by maintaining leaf relative water content, reducing hydrogen peroxide levels, & upregulating the expression of SlDREB1, SlDREB2, & SlDELLA genes.
Glycine max L.	GmNAC085	Drought	Overexpression in *Arabidopsis thaliana* enhances stress tolerance by reducing transpiration & cell membrane damage while activating drought-responsive gene expression. cell membrane damage while activating drought-responsive gene expression activating drought-responsive gene expression activating drought-responsive
Oryza sativa L.	OsICE1 & OsICE2	Cold	Overexpression in *Arabidopsis thaliana* enhances stress tolerance by upregulating cold-responsive genes. Overexpression in *Arabidopsis thaliana* enhances stress tolerance
Papaver somniferum	PsAP2	Abiotic / biotic	Overexpression in Nicotiana tabacum increases tolerance to both abiotic & biotic stresses tolerance to both abiotic & biotic stresses..

continued on next page

Table 2.2 continued from previous page

Plant System	**Transcriptions Factors**	**Stress**	**Function**
Saccharum officinarum var. Co740	SoMYB18	Salinity / drought	Expression in tobacco enhances stress tolerance by increasing chlorophyll content, antioxidant activity, & proline levels while decreasing malondialdehyde accumulation.
Solanum lycopersicum L	SlERF84	Drought / salt	Overexpression in *Arabidopsis thaliana* enhances stress tolerance by improving ROS-scavenging capacity.
S. lycopersicum L.	SlWRKY3	Salinity	Overexpression in Solanum lycopersicum L. enhances stress tolerance by increasing biomass & photosynthesis, along with higher accumulation of K+ & Ca2+ & reduced Na+ levels in leaves.
S. lycopersicum L.	SlNAM1	Chilling	Overexpression in tobacco enhances stress tolerance by improving germination, photosynthesis, & osmolyte content, while reducing wilting & ROS levels in transgenic plants.
S. lycopersicum L.	SlWRKY39	Salt / drought / biotic	Overexpression in Solanum lycopersicum L. enhances stress tolerance by upregulating stress-responsive gene expression.
S. tuberosum L. cv. Spunta	StDREB2	Drought	Overexpression in *Gossypium barbadense L.* enhances stress tolerance by increasing biomass, boll number, relative water content, soluble sugars, soluble protein, chlorophyll, proline, gas exchange parameters, & antioxidant enzyme activities, while reducing malondialdehyde, hydrogen peroxide, & superoxide anion levels.
Sorghum bicolor (Btx623)		Drought	Expression in *Arabidopsis thaliana* enhances tolerance by increasing antioxidant activity & proline content while decreasing malondialdehyde levels.

continued on next page

Table 2.2 continued from previous page

Plant System	**Transcriptions Factors**	**Stress**	**Function**
Triticum aestivum L.		Drought	Overexpression in transgenic wheat enhances stress tolerance by increasing survival rate, chlorophyll content, soluble sugar, & proline levels.
T. aestivem L	TaNAC47	Drought / salt / freezing	Overexpression in *Arabidopsis thaliana* enhances stress tolerance by modulating stress-responsive genes & altering multiple physiological parameters.
T. aestivum L.	TaWRKY93	Salinity / drought / low temperature	Overexpression in *Arabidopsis thaliana* improves tolerance to various stresses by preserving membrane stability, promoting osmotic balance, & modulating the expression of stress-related genes.
Zea mays (X178)	ZmWRKY106	Drought / heat	Overexpression in *Arabidopsis thaliana* enhances stress tolerance by strengthening the antioxidant system & regulating stress-related gene expression through the ABA signaling pathway.

2.6.11 Transcription Factors and Secondary Metabolism

Secondary metabolism in plants involves the production of compounds like alkaloids, terpenoids, and phenolics, which are not essential for basic survival but play crucial roles in defense, adaptation, and interaction with the environment. These secondary metabolites are also valuable for human use due to their medicinal, nutritional, and industrial properties. The biosynthesis of these compounds is tightly regulated by transcription factors (TFs), which control the expression of genes involved in various metabolic pathways [105].

Several TF families, including *AP2/ERF*, *bHLH*, *bZIP*, zinc finger proteins, MYB, and WRKY, are central to this regulation. These factors act as key switches, turning on or off specific genes that direct the production of secondary metabolites

based on developmental needs or environmental stimuli [106]. MYB transcription factors are widely recognized for their role in regulating the biosynthesis of compounds such as anthocyanins, which give plants their red, purple, and blue pigmentation; flavonoids, which function as antioxidants and protective agents; glucosinolates, important for defense; and other phenylpropanoids and lignins that contribute to structural integrity [107].

Similarly, AP2/ERF transcription factors have been identified as important regulators in multiple plant species, where they influence the production of specialized compounds such as artemisinin in Artemisia, taxol in Taxus, and terpenoid indole alkaloids in Catharanthus roseus. They also contribute to the regulation of metabolic pathways in commonly studied plants like tobacco and tomato, impacting flavor, aroma, and stress-related responses [108].

WRKY transcription factors also play a significant role in regulating secondary metabolism. They are involved in controlling the biosynthesis of phenolic compounds, various alkaloids, and terpenes. In specific plant systems, WRKYs have been shown to regulate the production of complex alkaloids like terpenoid indole alkaloids in *Catharanthus roseus* and benzylisoquinoline alkaloids in *Coptis japonica* [109].

TABLE 2.3: Examples of transcription factors in secondary metabolism [110].

Plant System	**Transcription Factors**	**Secondary Metabolites**
Arabidopsis thaliana	AtPAP1	Overexpression in *Leonurus sibiricus* L. enhances the production of key phenolic acids such as chlorogenic, neochlorogenic, ferulic, caffeic, and p-coumaric acids.
Artemisia annua	AabHLH112	Overexpression in *Artemisia annua* enhances the production of artemisinin.
Catharanthus roseus	CrERF5	Overexpression enhances the content of anhydrovinblastine, vinblastine, ajmalicine, vindoline, and catharanthine.
C. roseus	CrBIS2	Overexpression in *Catharanthus roseus* enhances the production of the anticancer alkaloids vincristine and vinblastine.

continued on next page

Table 2.3 continued from previous page

Plant System	Transcription Factors	Secondary Metabolites
Dendrobium officinale	DobHLH4	Overexpressing the DobHLH4 transcription factor in *Dendrobium officinale* petals boosts linalool production by activating key genes in its biosynthesis.
Nicotiana tabacum L	NtPMT1	Overexpression in *Nicotiana tabacum* L. increases nicotine production.
Ophiorrhiza pumila	OpWRKY2	Overexpression in *Ophiorrhiza pumila* leads to increased camptothecin production.
O. pumila	OpWRKY3	Overexpression in *Ophiorrhiza pumila* enhances the production of camptothecin.
Salvia miltiorrhiza	SmbZIP1	Overexpression in *Salvia miltiorrhiza* stimulates the biosynthesis of phenolic acids and tanshinones.
Salvia miltiorrhiza	SmbHLH148	Overexpression in *Salvia miltiorrhiza* enhances the production of caffeic acid, rosmarinic acid, salvianolic acid B, dihydrotanshinone I, cryptotanshinone, and tanshinone.

2.7 Molecular Basis of Floral Development in Plants

2.7.1 ABCDE Model of Flower Development

The ABCDE model of flower development is a foundational framework in plant developmental genetics. It describes how combinatorial expression of specific transcription factors determines the identity of floral organs in the concentric whorls of a typical angiosperm flower: sepals, petals, stamens, carpels, and ovules. This model emerged from classical genetic studies in *Arabidopsis thaliana* and *Antirrhinum majus*, where homeotic mutants plants in which one organ type is replaced by another helped delineate the roles of different gene classes [111].

In the ABCDE model:

a) A-class genes specify sepals (whorl 1),

b) A + B-class specify petals (whorl 2),

c) B + C-class specify stamens (whorl 3),

d) C-class alone specifies carpels (whorl 4),

e) D-class genes are involved in ovule identity, and

f) E-class genes, often *SEPALLATA*-like *MADS-box* genes, function as co-factors, forming quartets with ABCD-class proteins to stabilize their function [92].

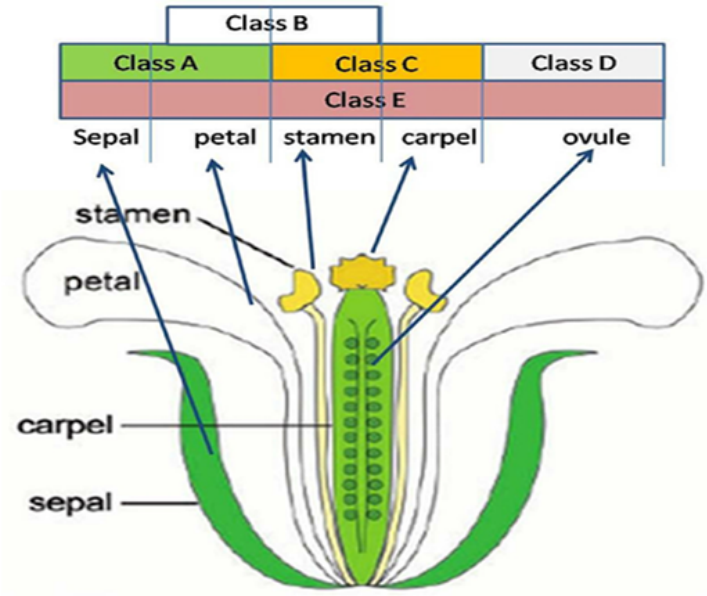

FIGURE 2.7: Representation of class A, B, C, D, and E genes with their respective tissue in model flower [111].

MADS-box transcription factors constitute the majority of ABCDE genes. These proteins contain a conserved *MADS* (*MCM1-AGAMOUS-DEFICIENS-SRF*) domain responsible for DNA binding and dimerization. Functional specificity arises from the formation of tetrameric complexes, often referred to as the "floral quartets," that bind to CArG-box motifs in the promoters of downstream target genes [112].

In addition to offering a mechanistic perspective on floral development, the ABCDE model acts as a framework for deciphering gene function across a variety of species. Homologous ABCDE genes have been found in polyploid crops like *Gossypium hirsutum*, and their expression patterns point to conserved roles in floral organ specification. However, layers of regulatory complexity are introduced by the presence of multiple homeologs as a result of polyploidy, which may permit the subfunctionalization or neofunctionalization of gene duplicates [113].

2.7.2 Key Regulatory Genes in Flower Development

Several key regulatory genes act upstream or in parallel with ABCDE-class genes to initiate and maintain floral development. These include:

a) *LEAFY* (*LFY*): A master regulator that promotes the transition from vegetative to floral meristem. It directly activates A-class genes such as *APETALA1* and plays a critical role in initiating the floral developmental program [95].

b) *APETALA1* (*AP1*) and *CAULIFLOWER* (*CAL*): Function redundantly to specify floral meristem identity and are essential for the proper initiation of the floral organ identity cascade.

c) *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) and *AGL24*: Act in the flowering time pathway, integrating environmental and endogenous cues such as photoperiod, vernalization, and gibberellin signaling to time floral induction.

d) *WUSCHEL* (*WUS*) and *CLAVATA3* (*CLV3*): Involved in meristem maintenance and size control. They ensure a sufficient pool of undifferentiated cells for organ initiation while preventing over proliferation [114].

Complex feedback loops and feed-forward mechanisms are frequently used to interact with these regulatory genes. Both transcriptional and post-transcriptional

mechanisms, such as protein-protein interactions, chromatin remodeling, and microRNA regulation, tightly regulate their expression. For example, to limit carpel identity to inner floral whorls and maintain organ patterning fidelity, *APETALA2*'s suppression of *AGAMOUS* in outer floral whorls is essential. Through transcriptomic and phylogenomic analyses, orthologs of these important regulatory genes have been found in the context of *Gossypium hirsutum*. Their conserved roles across angiosperms are further supported by the correlation between their expression patterns during flower development stages and changes in meristem identity and organogenesis [115].

2.7.3 Role of MADS-box Transcription Factors

MADS-box genes are among the most studied transcription factors in plant developmental biology due to their central roles in regulating flower development, seed formation, and fruit ripening. They are defined by a conserved ∼58 amino acid MADS domain and are classified into two major groups in plants: type I (SRF-like) and type II (MIKC-type).

Type II MIKC-type genes, which include the majority of ABCDE floral genes, are further subdivided based on their additional I (intervening), K (keratin-like), and C (C-terminal) domains [116].

These genes operate through the formation of multimeric complexes, which bind specific DNA motifs to activate or repress downstream gene expression. Structural studies have shown that the K-domain facilitates dimer and tetramer formation, critical for functional specificity. The C-terminal region, while variable, often contains activation domains or motifs for protein-protein interactions.

Because of polyploidy, MADS-box genes are widely duplicated in *Gossypium hirsutum*. According to phylogenetic and synteny analyses, the *SAP* genes in this family share exon-intron structures and conserved motifs with their orthologs in rice, tomato, and *Arabidopsis*. However, floral morphology, especially petal and

stamen identity, can be greatly influenced by minute variations in domain composition or expression timing. Additionally, MADS-box transcription factors react to environmental cues like temperature, photoperiod, and nutritional status.

A crucial characteristic for crop adaptation, this responsiveness enables plants to modify organ development and flowering timing in response to changing circumstances. For instance, in cotton, certain MADS-box genes exhibit altered expression in response to heat stress or drought, indicating a potential function in preserving reproductive success in challenging circumstances [117].

Importantly, mutations in specific MADS-box genes often result in homeotic transformations. For example, loss-of-function mutations in B-class genes such as *PISTILLATA* or *APETALA3* lead to the replacement of petals with sepals and stamens with carpels. These phenotypes, observed across multiple species, underscore the conserved and essential roles of MADS-box proteins in defining floral organ identity [118].

2.8 The *Sterile Apetala* (*SAP*) Gene Family

2.8.1 Historical Discovery and Characterization

Mutations in the *SAP* gene caused *Arabidopsis thaliana* flowers to produce sterile structures and lack petals, which led to the initial identification of the *STERILE APETALA* (*SAP*) gene family through mutant analyses . The significance of *SAP* as a crucial regulator of floral organ identity, particularly in the development of petals and stamens, was highlighted by these homeotic changes. Early research identified the *SAP* locus as belonging to the MADS-box transcription factor family using positional cloning and genetic mapping. This family is distinguished by the highly conserved MADS domain, which is necessary for dimerization and DNA binding and enables these proteins to act as master regulators in floral development.

The discovery of *SAP* genes in *Arabidopsis* provided a foundation for exploring their orthologs and paralogs across diverse angiosperms. Comparative genomics has demonstrated that *SAP*-like genes are broadly conserved across flowering plants, reflecting their essential roles in maintaining the structural integrity and reproductive success of flowers. Beyond *Arabidopsis*, studies in crops such as tomato, rice, and maize have identified homologous genes with similar functions in floral development. This conservation has allowed researchers to leverage model systems to predict gene functions in economically important species like cotton [119].

2.8.2 Structure and Domains of *SAP* Genes

SAP genes, like other MADS-box transcription factors, possess a modular structure comprising several conserved domains:

2.8.2.1 MADS Domain (M)

Approximately 58 amino acids long, responsible for DNA binding to CArG-box motifs in target gene promoters. This domain also facilitates dimerization with other MADS proteins .

2.8.2.2 Intervening Domain (I)

Provides specificity in dimer formation, contributing to selective interactions with other MADS-box partners.

2.8.2.3 Keratin-like Domain (K)

Characterized by coiled-coil motifs that stabilize dimer and tetramer formations. This domain is critical for forming higher-order complexes that regulate gene expression precisely.

2.8.2.4 C-terminal Domain (C)

Often variable among family members but can contain transcriptional activation motifs or interaction sites for co-regulators [120].

The functional diversity of *SAP* genes is supported by this conserved domain architecture. *SAP* proteins can control a variety of target gene sets involved in floral organ specification, meristem identity, and reproductive development by forming complexes with other MADS-box proteins through combinatorial interaction. The ways in which these domains mediate DNA binding and protein-protein interactions have been thoroughly elucidated by structural investigations, such as X-ray crystallography and molecular modeling. Understanding how even minormutations can impair *SAP* function and result in striking phenotypic effects requires this kind of information [120].

2.8.3 Functional Role in Model Plants

In *Arabidopsis*, *SAP* mutants exhibit striking phenotypes characterized by sterile, apetalous flowers. These phenotypes result from the failure to correctly specify petal and stamen identities in the second and third whorls of the flower. Instead, these whorls often develop into sepaloid or carpeloid structures, reflecting a loss of B-class gene activity [121].

Functional analyses using transgenic complementation and expression studies have confirmed *SAP*'s role as a positive regulator of B-class genes such as *APETALA3* (AP3) and *PISTILLATA* (PI). *SAP* is thought to act upstream or in complex with these genes, stabilizing their expression domains and ensuring the correct spatial patterning of petals and stamens [122].

SAP transcripts are highly localized to developing floral meristems, especially during the early stages of organ initiation, according to expression profiling. Its function in determining organ identity is consistent with its temporal and spatial specificity. Further highlighting *SAP*'s function as a master regulator, ectopic

expression studies have shown that misregulation of *SAP* can result in petaloid structures in organs that are normally non-petalous.

These functional discoveries from *Arabidopsis* offer a model for examining the functions of *SAP* genes in different organisms. Although polyploidy and gene duplication may introduce extra regulatory layers, the conservation of floral regulatory networks indicates that comparable mechanisms may function in *Gossypium hirsutum* [123].

2.8.4 *SAP* Orthologs and Paralogs in Other Species

Beyond *Arabidopsis*, *SAP* orthologs and paralogs have been identified in numerous plant species, reflecting the evolutionary conservation of their roles in floral development. Comparative genomics and phylogenetic analyses have revealed that *SAP*-like genes are present in most angiosperm lineages, often retaining their characteristic domain structures and regulatory functions [124].

For instance, TM6, a B-class MADS-box gene that is closely related to *SAP*, controls the development of the petal and stamen in tomatoes (*Solanum lycopersicum*). Similar to the sap phenotype of Arabidopsis, loss-of-function mutants display diminished or deformed petals and stamens. Likewise, MADS-box genes like OsMADS16 have comparable functions in determining the identity of lodicules and stamens in rice (*Oryza sativa*). According to phylogenetic analyses, *SAP* genes frequently create unique clades within the MADS-box gene family, which is indicative of functional diversification and lineage-specific duplications. In polyploid species, where several homeologous copies may develop novel expression patterns or regulatory interactions a process referred to as sub functionalization or neo functionalization these duplications are especially noticeable [125].

Several putative *SAP* orthologs and paralogs have been found in *Gossypium hirsutum* through bioinformatics analyses, indicating a complex gene family structure influenced by its allopolyploid genome. Predicting functional conservation and divergence requires an understanding of these relationships. The evolutionary

forces that have influenced cotton's floral development strategies can be revealed through comparative studies of *SAP* gene sequences and expression profiles among *Gossypium* species [126].

These evolutionary viewpoints have applications in addition to advancing basic knowledge. Breeders and geneticists can manipulate particular *SAP* gene copies to maximize floral traits for higher cotton yield and quality by identifying conserved functional domains and expression patterns.

2.9 *Sterile Apetala* Genes In *Gossypium* Species

2.9.1 Identification of *SAP* Genes in *Gossypium*

The identification of *STERILE APETALA* (*SAP*) genes in *Gossypium* species represents a crucial step in understanding the genetic regulation of floral development in cotton. Due to its allopolyploid nature, *Gossypium hirsutum* harbors duplicated gene copies derived from its A- and D-genome progenitors. This polyploid complexity can obscure straightforward gene identification, necessitating advanced bioinformatics and comparative genomics approaches [127].

Recent genome sequencing initiatives, such as those producing high-quality assemblies of *G. hirsutum, G. arboreum*, and *G. raimondii*, have enabled researchers to systematically mine for SAP homologs. Using BLAST-based homology searches, Hidden Markov Models (HMMs), and domain annotation pipelines like InterProScan and Pfam, scientists have identified multiple putative SAP gene candidates in *Gossypium genomes*.

Genome-wide studies have also employed transcriptomic data to validate the expression of *SAP* gene candidates in floral tissues. Such integrative approaches help distinguish functional genes from pseudogenes or non-expressed duplicates. Moreover, expression patterns observed in developing buds and floral organs provide crucial evidence for their roles in floral organ identity.

The identification of SAP genes in *Gossypium* lays the groundwork for functional validation studies. These efforts are essential for confirming whether the predicted homologs maintain conserved roles akin to those observed in model systems such as *Arabidopsis thaliana* or exhibit cotton-specific regulatory innovations arising from polyploid evolution [128].

2.9.2 Evolutionary Divergence and Gene Duplication Event

Gossypium hirsutum's evolutionary history is marked by polyploidization events that have driven extensive gene duplication and divergence. The allopolyploid genome of *G. hirsutum* contains two subgenomes (At and Dt) derived from ancestral A- and D-genome diploids. This history has resulted in homeologous gene pairs, including multiple copies of MADS-box and SAP genes.

Gene duplication provides raw material for evolutionary innovation. Duplicated genes can undergo:

a) Neo functionalization, where one copy acquires a new function;

b) Sub functionalization, where ancestral functions are partitioned between duplicates; or

c) Non functionalization, leading to pseudogenization [129]

In the context of *SAP* genes, duplication events may have facilitated the diversification of floral developmental programs in *Gossypium* species. For example, subtle differences in expression timing or tissue specificity between homeologs could enable fine-tuned regulation of petal and stamen development under varying environmental conditions.

Comparative genomic analyses across *Gossypium* species have begun to elucidate the evolutionary trajectories of SAP genes. Phylogenetic trees incorporating *SAP* orthologs and paralogs from *G. hirsutum, G. arboreum,* and *G. raimondii* reveal

patterns of conservation and divergence consistent with polyploidy-mediated evolution. Such studies help pinpoint candidate genes for functional assays and inform hypotheses about the adaptive significance of floral regulatory diversity in cotton [130].

2.9.3 Reported Functional Studies

Despite progress in identifying and annotating SAP genes in *Gossypium* genomes, functional validation studies remain limited. Experimental analyses directly linking SAP gene function to floral phenotypes in cotton are scarce, reflecting both the challenges of working with polyploid genomes and the relatively recent availability of genomic resources.

Nevertheless, some indirect evidence supports putative roles for *SAP* genes in *Gossypium hirsutum*. Transcriptome profiling has shown that predicted *SAP* homologs are differentially expressed during floral development stages, with peak expression often observed in floral meristems and developing petals and stamens. Such expression patterns mirror those of *SAP* genes in model plants, suggesting conserved roles in organ identity specification [131].

Furthermore, a possible connection between *SAP* activity and reproductive success has been suggested by comparative expression analyses between sterile and fertile cotton lines, which have suggested a differential regulation of *SAP* gene expression. These findings highlight the significance of seeking functional validation using gene editing or silencing techniques, even though they are correlative. New genetic technologies like Virus-Induced Gene Silencing (VIGS) and CRISPR-Cas9 present encouraging paths for the functional analysis of *SAP* genes in cotton. Researchers can directly evaluate the roles that particular *SAP* gene copies play in petal and stamen development, floral fertility, and overall yield by knocking out or down regulating those copies. Confirming predicted functions and guiding breeding strategies to maximize floral traits in *Gossypium hirsutum* will require such studies [132].

Beyond basic science, understanding *SAP* gene function has translational potential for cotton improvement. Manipulating floral development can influence traits such as boll set, pollinator attraction, and reproductive synchrony, all of which are important targets for breeding programs seeking to enhance yield, quality, and resilience under changing environmental conditions [133].

2.10 Bioinformatics Approaches and Pipelines

2.10.1 Databases and Genomic Resources for Cotton

Access to comprehensive and curated genomic resources is critical for the bioinformatics analysis of *SAP* genes in *Gossypium hirsutum*. One of the most valuable resources is CottonGen, an online database integrating diverse datasets:

a) Whole-genome assemblies of *G. hirsutum, G. arboreum,* and *G. raimondii.*

b) Annotated gene models with functional descriptions.

c) Transcriptomic datasets from different tissues and developmental stages.

d) Molecular markers and quantitative trait loci (QTL) relevant to breeding.

CottonGen simplifies access to crucial resources for researchers looking into *SAP* genes by providing user-friendly interfaces for BLAST searches, genome browsers, and data downloads. Gene locations, exon-intron structures, and nearby regulatory elements can be zcotton genomic data is also available in other public repositories such as NCBI, Ensembl Plants, and Phytozome, which provide complementary annotations and redundancy. By facilitating the cross-validation of gene models, these resources guarantee the accuracy of *SAP* gene identification [134].

2.10.2 Tools for Gene Structure Prediction

Accurate prediction of gene structure is essential for characterizing *SAP* genes, especially in polyploid species with extensive gene duplications. Popular tools include:

a) *AUGUSTUS*: *Ab initio* gene prediction software trained on plant genomes, capable of predicting exon-intron boundaries with high accuracy.

b) *FGENESH*: Another *ab initio* predictor, often used in combination with evidence-based approaches to refine gene models.

c) *BRAKER*: Combines RNA-Seq data with ab initio predictions, integrating expression evidence to improve accuracy [135].

These tools allow researchers to:

a) Confirm the integrity of predicted *SAP* gene models.

b) Distinguish functional genes from pseudogenes.

c) Analyze splice variants that might have tissue- or stage-specific roles.

For *G. hirsutum*, integrating these predictions with transcriptomic data (e.g., full-length Iso-Seq reads) ensures high-confidence annotations, essential for downstream functional analyses and comparative genomics [136].

2.10.3 Phylogenomic Methods

Phylogenomics provides a powerful framework for understanding the evolutionary history of *SAP* genes. Unlike traditional single-gene phylogenies, phylogenomic approaches:

a) Analyze large sets of orthologous and paralogous genes across species.

b) Incorporate whole-genome or transcriptome datasets.

c) Offer greater resolution in inferring duplication events and divergence times.

For *SAP* genes in cotton, phylogenomic analyses can:

a) Reveal whether homeologous gene pairs resulted from ancient whole-genome duplications or more recent polyploidization events.

b) Clarify orthologous relationships with *SAP* genes in model species such as *Arabidopsis* or tomato.

c) Support hypotheses about sub functionalization or neo functionalization following duplication.

Tools such as OrthoFinder, IQ-TREE, RAxML, and MrBayes are commonly used for phylogenomic reconstructions, while visualization packages like FigTree and iTOL enable intuitive interpretation of complex evolutionary histories [137].

2.10.4 Functional Annotation Tools

Accurate functional annotation of *SAP* genes relies on identifying conserved domains and motifs. Widely used tools include:

a) InterPro: An integrative platform combining multiple protein signature databases (Pfam, SMART, PROSITE, SUPERFAMILY). It provides domain annotations, GO terms, and pathway information.

b) Pfam: Specializes in protein family classification based on hidden Markov models, crucial for identifying the MADS-box domain and related motifs.

Functional annotation of *SAP* gene candidates confirms their identity as MADS-box transcription factors, distinguishing them from unrelated proteins. Beyond confirming the hallmark MADS domain, annotation tools can:

a) Reveal K-domains involved in protein-protein interactions.

b) Detect variable C-terminal regions that might confer novel regulatory functions.

c) Suggest potential post-translational modification sites impacting activity.

These analyses also support comparative genomics, enabling researchers to assess domain conservation across orthologs and paralogs in different species [137].

2.10.5 Integrative Approaches

Modern bioinformatics emphasizes the integration of multiple "omics" layers to achieve a systems-level understanding of gene function. For *SAP* genes in *Gossypium hirsutum*, such integration may include:

a) *Genomics*: High-quality reference genomes enabling precise gene localization and annotation.

b) *Transcriptomics*: RNA-Seq and qPCR data providing expression profiles across tissues, stages, and environmental conditions.

c) *Proteomics*: Mass spectrometry data confirming protein presence, post-translational modifications, and interaction partners.

d) *Epigenomics*: DNA methylation and histone modification profiles influencing gene regulation [138].

Integrating these datasets can:

a) Reveal dynamic regulatory networks controlling *SAP* gene expression.

b) Identify co-expression modules linking *SAP* genes to other floral regulators.

c) Highlight environmental or developmental cues impacting *SAP* gene activity.

For example, integrating RNA-Seq data with promoter methylation maps can explain expression divergence between homeologous *SAP* genes in polyploid cotton. Such insights are invaluable for functional prediction and breeding applications.

Advanced bioinformatics pipelines now incorporate machine learning and AI-based approaches for multi-omics integration, offering new avenues for hypothesis generation and experimental prioritization [139].

Chapter 3

Methodology

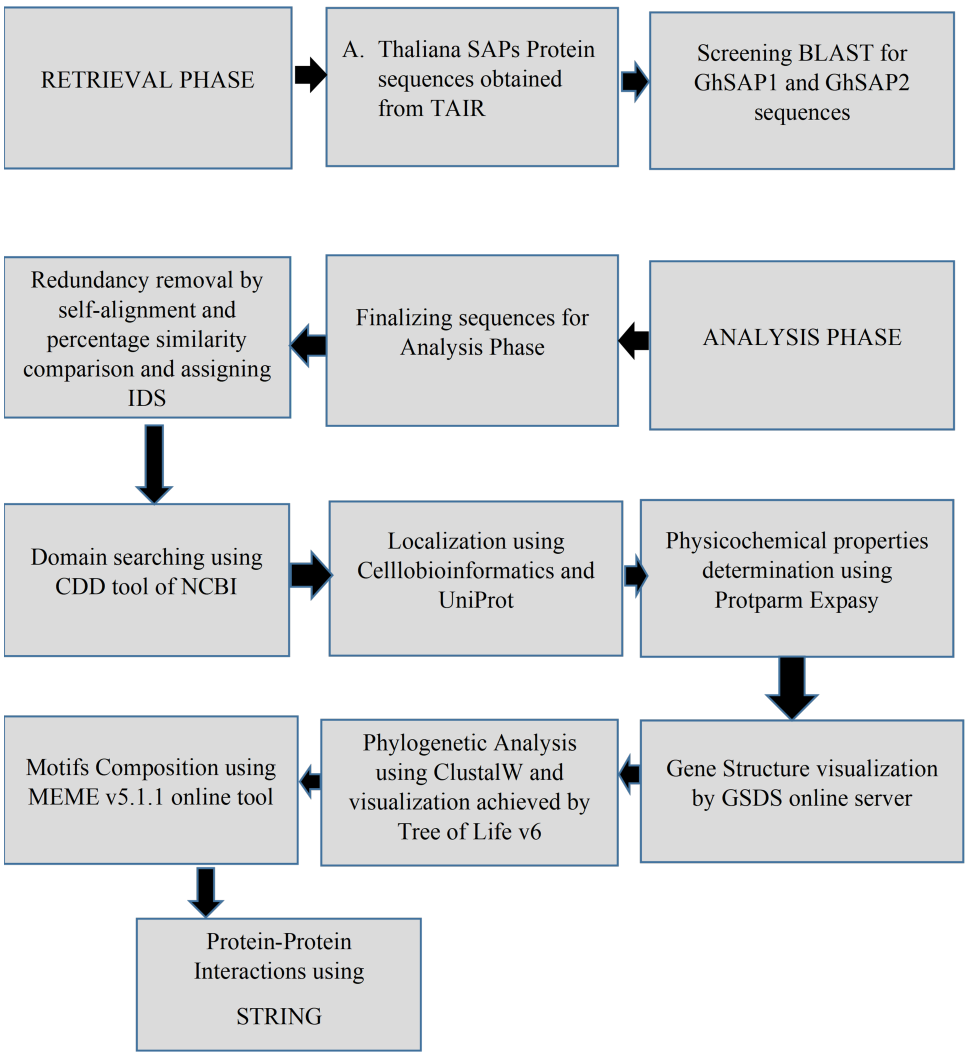

FIGURE 3.1: Flowchart of Methodology

3.1 Screening of Genome and Transcription Factors Databases

The *Arabidopsis* genome resource TAIR (http://arabidopsis.org) was utilized to get the entire gene, amino acid, and coding sequences for each member of the *Arabidopsis thaliana SAP* gene family. The recognition of possible *Arabidopsis thaliana SAPs* was examined using three protein databases: NCBI (www.ncbi.nlm.nih.gov/), TAIR (https://www.arabidopsis.org/), and TFDB (planttfdb.gao-lab.org).

3.2 GhSAPs Protein Sequences Retrieval

The *AtSAP* protein sequences were employed as query sequences in two database i.e NCBI and TFDB to identify protein sequences for *SAP GhSAP*. Additionally, TAIR was screened using accession numbers to retrieve *GhSAP* protein sequences.

The sequences were obtained in FASTA format and stored for use in subsequent sampling and analyses. Two databases were screened for the identification and characterization of any putative *GhSAPs* and all *GhSAPs* sequences were downloaded and aligned for further analysis.

3.3 Removal of Redundant Sequences of *GhSAPs*

The protein sequences of *GhSAP* were aligned to eliminate redundancy. Repetition among the accession numbers of each sequence, obtained from different databases such as NCBI or TFDB, was removed. The sequences were rearranged, and matching was performed to identify identical sequences from different databases. This process ensures that no sequence is repeated, and the samples are specific and accurate.

3.4 Allocation of Lab IDs and Establishment of Sequence Similarity of *GhSAP1* and *GhSAP2* Protein Sequences

At first, Lab-IDs were given to each sequence after the removal of redundant, spliced and incomplete sequences. Secondly, to determine the sequence similarities between the *GhSAPs* protein sequences a comparison table was generated. Sequences with similarity below the threshold (80%) would not be taken into consideration.

3.5 Physicochemical Properties and Conserved Domains in *AtSAP* and *GhSAP1* and *GhSAP2*

Potential *GhSAPs* were chosen from the retrieved sequences based on the conserved domains. The chosen genes had both F-BOX domain and WD40 repeat-like domains. Protparam Expasy (web.expasy.org/protparam/) was used to examine the physical and chemical characteristics of *AtSAP*, *GhSAP1* and *GhSAP2* including their protein molecular weight (MW), hydropathicity (GRAVY), and theoretical isoelectric point (Pl), while their subcellular localizations was predicted via CELLO (http://cello.life.nctu.edu.tw/).

3.6 Motif Composition in *GhSAPs* Gene Family

The presence of consensus motif was determined using online tool called Multiple Em for Motif Elicitation or MEME v5.1 (https://meme-suite.org/meme/ tools/meme). It is the most accurate online tool for motif elicitation. All parameter setting were kept at default settings with an exception of motif threshold, which

was kept at 9 to ensure specificity and precision. The identification of consensus motif will provide evidence of local regions that are similar between scale and study organism.

3.7 Phylogenetic Analysis of *GhSAPs*

To generate the phylogenetic tree for *GhSAP* genes, in which different species of *Gossypium hirsutum* and *Arabidopsis thaliana*, were aligned through Clustal Omega. Interactive Tree of Life software v6 (https://itol.embl.de/) was used to create the phylogenetic tree of GOSSS with the selected plants to illustrate the evolutionary link between *GhSAP1* and *GhSAP2* with other plants.

3.8 Gene structure and Motif Composition in *GhSAPs* Gene Family

The coding and full length gene sequences of *GhSAPs* was used to examine gene structural components using GSDS online server (Gene Structure Display Server: (http://gsds.cbi.pku.edu.en/). The introns, exons, and un-translated regions were identified. This tools assisted in determining exons, intrans and untranslated regions (UTRs) present in the sequence. This helps in comparing gene length between two families.

3.9 Protein-Protein Interaction of *GhSAPs*

The cellular proteins in interaction with *GhSAPs* and *AtSAP* were predicted using an online tool called STRING (https://string-db.org/), and their figures along with the description, was exported from the tool for comparative study and future reference. The interacting proteins of *GhSAPs* were compared with those of AtSAPs to identify any functional homology. The colour lines show known and

predicted interactions, the pink and blue colour shows known interactions and dark green, blue and red shows predicted interactions. These balls that are in different colours shows nodes while the coloured lines are the edges means protein sequences.

3.10 Promoter Analysis

Gene regulatory elements in *GhSAPs* promoter regions were identified by analyzing upstream promoter regions (2000 bps) obtained from an online database, Plant Ensembl (https://plants.ensembl.org/). Cis-regulatory elements were examined in promoter regions using a database (https://bioinformatics.psb.ugent.be/plantcare/).

Chapter 4

Results

4.1 Genome - Wide Identification and Analysis of *SAP* Gene in *Gossypium hirsutum*

In the current research, the full-length protein-coding sequences of the *SAP* gene from the *Arabidopsis thaliana SAP* gene family were retrieved by entering the accession number of *SAP* on the TAIR database (TAIR: http://arabidopsis.org/). There were 8 variants of the *SAP* gene and the longest sequence was chosen. Then the longest sequence was Protein Blast in the NCBI database (www.ncbi.nlm.nih.gov/genome/gdv/). The results were downloaded in description CSV format in an Excel file.

All the results were saved with name, accession numbers, and lab ids in an Excel file. They were self-aligned to make a percentage similarity table of all downloaded sequences After the similarity table, all the spliced variants, repeated/redundant sequences, and short or incomplete fragments were excluded from retrieved sequences simultaneously validated through conserved domain identification, and selected 2 sequences of *Goosypium hirsutum* were downloaded in the FASTA format and a new Excel sheet of selected sequences with their lab ids, gene names, and accession numbers was made. *SAP*s in the *Gossyoium hirsutum* genome were identified using the *Arabidopsis thaliana SAP* protein sequences as well as the

pfam accessions of F Box (IPR036047) and WD40 Repeat –like (IPR036322) as queries.

TABLE 4.1: Sequence similarity among protein sequences of GhSAPs

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
S1	100	96.4	78.16	78.7	100	77.84	78.92	55.43
S2	96.4	100	78.03	78.57	95.97	77.69	78.76	54.17
S3	78.16	78.03	100	94.81	78.11	99.73	97.57	55.41
S4	78.97	78.84	94.81	100	79.19	97.53	99.73	54.93
S5	100	95.97	78.11	79.19	100	77.84	78.92	57.88
S6	77.84	77.69	99.73	97.57	77.84	100	97.98	58.68
S7	78.92	78.76	97.57	99.73	77.92	97.98	100	58.99
S8	55.48	54.8	55.1	53.59	57.88	58.68	58.99	100

4.2 Physicochemical Properties and Conserved Domains of *Gossypium hirsutum*

2 *GhSAPs* were identified that contained both F-Box and WD40 repeat-like domains and were labelled from 1 to 8 with respect to chromosome numbers. Accession numbers of same or redundant sequences found in selected databases are enlisted. Potential *GhSAPs* were chosen from the retrieved sequences based on the conserved domains discovered database by the CDD (www.ncbi.nlm.nih.gov). All of these sequences and query sequences were pasted in CD search in CCD and submitted.

The result is summarized in (Table 4.2) which includes query, hit type, position, E value, bit-score, Accession, and short name of both plants. Whereas, the physical and chemical properties of *A. thaliana* and *Gossypim histurum SAP* gene families including protein molecular weight (MW), hydropathicity (GRAVY), and theoretical isoelectric point (pl) values of both plants were examined in the Expasy Prot Param tool and are summarized in (Table 4.3). The gene lengths, protein lengths, and molecular weights, pl and GRAVY values of both plants were close to each other. The values of both plants showed negative GRAVY values which showed

as hydrophilic proteins. The PI and GRAVY values of both plants are close to each other. While the sub-cellular localizations were examined which proposed *A. thaliana* and *G. hirsutum* of *SAPs* to be localized in Chloroplast.

AtSAP, *GhSAP1* and *GhSAP2* were found to have average protein lengths of 446,453 and 463 amino acids, respectively. However, there was significant variation between the gene lengths of AtSAPs and *GhSAPs*. *AtSAPs* and *GhSAPs* showed PI values below 7, indicating that they are acidic proteins.

TABLE 4.2: Conserved Domains of *SAP* gene families of *GhSAP1* and *GhSAP2*

Organism	**Query**	**Hit Type**	**Position**	**E value**	**Bit Score**	**Accession**	**Short Name**
Gossypium hirsutum	*GhSAP1*	Specific	162-449	9.3749e-08	54.1668	COG2319	WD40
	GhSAP2	Super family	166-459	2.38449e-05	46.174	Cl29593	W440 Superfamily

TABLE 4.3: Physical and chemical properties of SAP gene families of *Arabidopsis thaliana* and Gossypium hirsutum

Attribute	**Arabidopsis**	**Cotton - AtSAP2**	**Cotton - AtSAP4**
Plant	Arabidopsis	Cotton	Cotton
Gene name	AtSAP	AtSAP2	AtSAP4
Chr	5	A07	D07
Position	13936086–13940867	91197501–1202260	35653307–35658045
Gene length	4782	4760	3739
Protein length	446	453	463
Molecular weight	49306.46	501199.1	50642.58
Isoelectric point	6.74	5.93	6.49
GRAVY	-0.145	-0.175	-0.224
Localization	Plasma membrane	Chloroplast	Extracellular

4.3 Gene Structure Determination

The gene and coding sequences of *AtSAPs* and *GhSAPs* were used to analyze their structural features which included the identification of exons, Introns and

untranslated regions (UTRs).This was done using an online server called GSDS or Gene Structure Display Server.

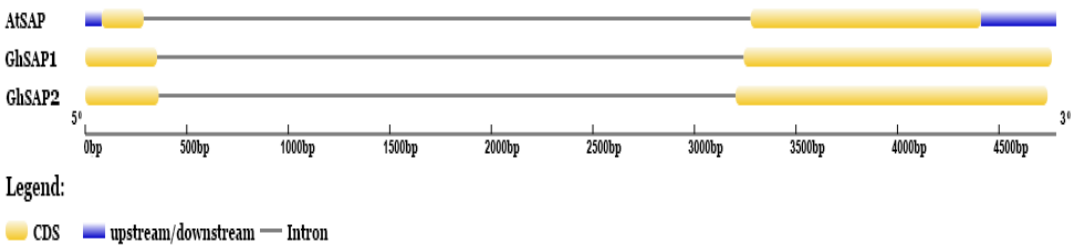

FIGURE 4.1: Displays the result of Gene Structure Determination

4.4 Displays the Result of Gene Structure Determination

The interacting *SAP* proteins networks were predicted online through STRING (Table 4.4). All the Gh*SAP* proteins were suggested to interact with related genes. The interaction network shows that the *SAP* (*Sterile Apetala*) protein in *Arabidopsis thaliana* is closely connected with a variety of other proteins, suggesting it plays a central role in multiple biological processes.

Notably, *SAP* interacts with members of the *TIFY* protein family (TIFY4A and TIFY4B), which are known for their role in plant stress responses and development.

It also links with proteins like OPT5, SCPL12, and IQD33, which are involved in transport, metabolism, and cellular signaling. These connections hint that *SAP* may function not only in floral development, as previously known, but also in broader pathways related to plant growth, defense, and adaptation.

The strong associations seen here suggest that *SAP* could be a key regulatory hub in the plant's response to its environment.

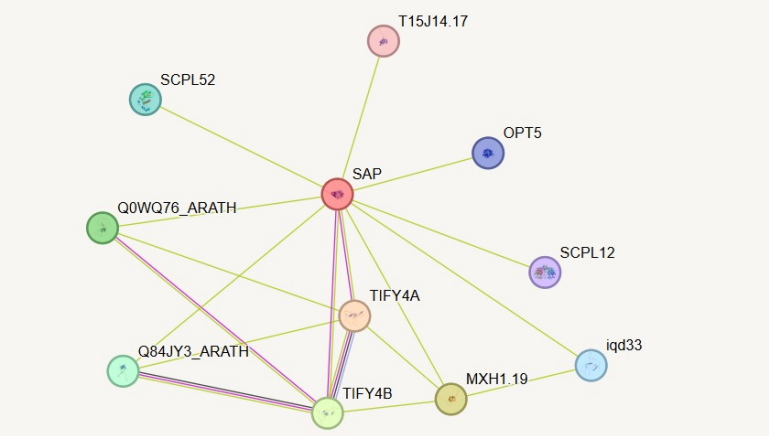

FIGURE 4.2: *AtSAP* Protein-Protein Interaction

Predicted Functional Partners

TABLE 4.4: AtSAP Protein Protein Interaction

TIFY4A	Protein TIFY 4A; Regulates the arrest of dispersed meristematic cells during lamina development.
MXH1.19	Beta-galactosidase related protein.
TIFY4B	Protein TIFY 4B; Regulates the arrest of dispersed meristematic cells during lamina development.
Q0WQ76_ARATH	Uncharacterized protein At3g24150.
Q84JY3_ARATH	Histone acetyltransferase.
SCPL52	Putative serine carboxypeptidase-like 52.
Iqd33	IQ-Domain 33
SCPL12	Serine carboxypeptidase-like 12; Probable carboxypeptidase.
T15J14.17	Plant basic secretory protein (BSP) family protein.

Protein-Protein Interaction of *GhSAP1*

In (Fig4.2) a strong web of interactions among a group of proteins, with LOC10789-0429 appearing at the center. The dense connections among these proteins represented by blue lines suggest they likely work closely together in the same biological pathway or process. Several nodes, such as LOC107959578, LOC107962082, and LOC107944107, also serve as highly connected hubs, indicating they may play key roles in coordinating or regulating shared functions. This kind of tight-knit clustering often points to a protein complex or a coordinated response system, such as involvement in stress signaling, development, or metabolic regulation. Overall,

the image highlights a cooperative network, where each protein supports or influences the activity of others, suggesting a highly integrated functional module in the plant.

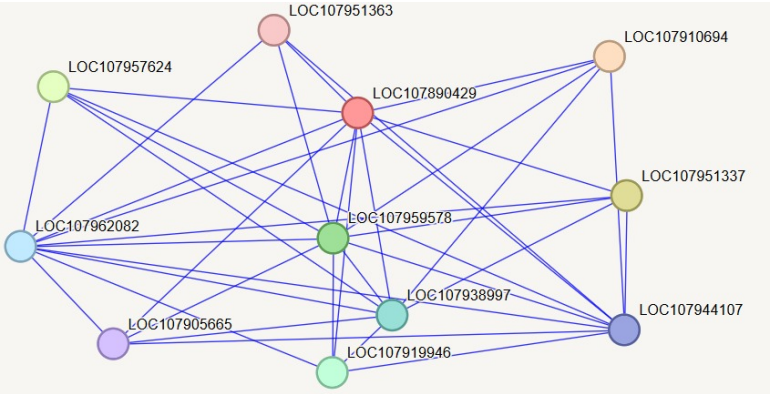

FIGURE 4.3: *GhSAP1* Protein Protein Interaction

TABLE 4.5: *GhSAP1* Protein-Protein Interaction

LOC107910694	Uncharacterized protein LOC107910694.
LOC107951337	Uncharacterized protein LOC107951337.
LOC107957624	Uncharacterized protein LOC107957624.
LOC107959578	Uncharacterized protein At4g38062-like.
LOC107919946	Uncharacterized protein At4g26450-like isoform X1
LOC107938997	Protein XRI1-like isoform X1.
LOC107962082	CASP-like protein; Belongs to the Casparian strip membrane proteins (CASP) family.
LOC107944107	Protein XRI1-like.
LOC107905665	Uncharacterized protein LOC107905665.
LOC107951363	Uncharacterized protein At4g26450-like isoform X1.

Protein-Protein Interaction of *GhSAP2*

This interaction map shows a close-knit group of proteins working together, with LOC107956754 appearing as one of the main coordinators in the center. It's strongly connected to several other proteins like LOC107959578 and LOC107962082, suggesting they may be involved in the same biological role almost like teammates on a shared project. The dense network of links between all these proteins shows they're not working alone; instead, they likely depend on each other to carry out important functions in the plant, possibly in response to stress or during growth.

This kind of teamwork at the molecular level is a strong sign that these proteins are part of a larger, well-organized system keeping the plant functioning properly.

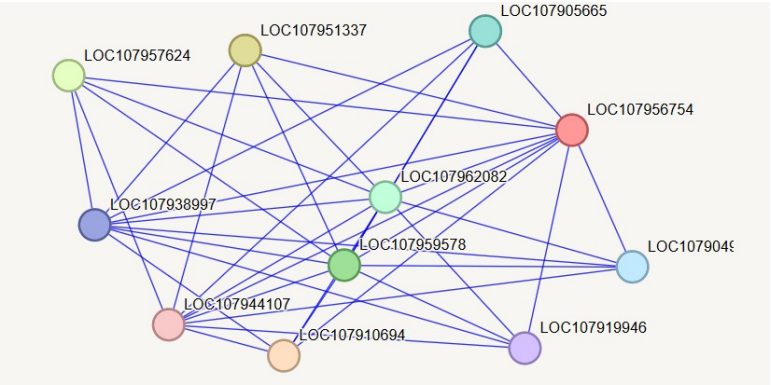

FIGURE 4.4: *GhSAP2* Protein-Protein Interactions

TABLE 4.6: GhSAP2 Protein-Protein Interaction

LOC107910694	Uncharacterized protein LOC107910694.
LOC107951337	Uncharacterized protein LOC107951337
LOC107957624	Uncharacterized protein LOC107957624
LOC107959578	Uncharacterized protein At4g38062-like.
LOC107962082	CASP-like protein; Belongs to the Casparian strip membrane proteins (CASP) family.
LOC107905665	Uncharacterized protein LOC107905665.
LOC107904976	Uncharacterized protein LOC107904976.
LOC107938997	Protein XRI1-like isoform X1
LOC107919946	Uncharacterized protein At4g26450-like isoform X1.
LOC107944107	Protein XRI1-like.

4.5 Consensus Motifs Composition in ATSAPs and GhSAPs

For identification of consensus motifs, *AtSAPs* and *GhSAPs* protein sequences were selected for 10 consensus motifs as shown in figure 4.5. All the sequences featured motifs that were remarkably conserved in the proteins of *Arabidopsis thaliana* and Gossypium histurum.

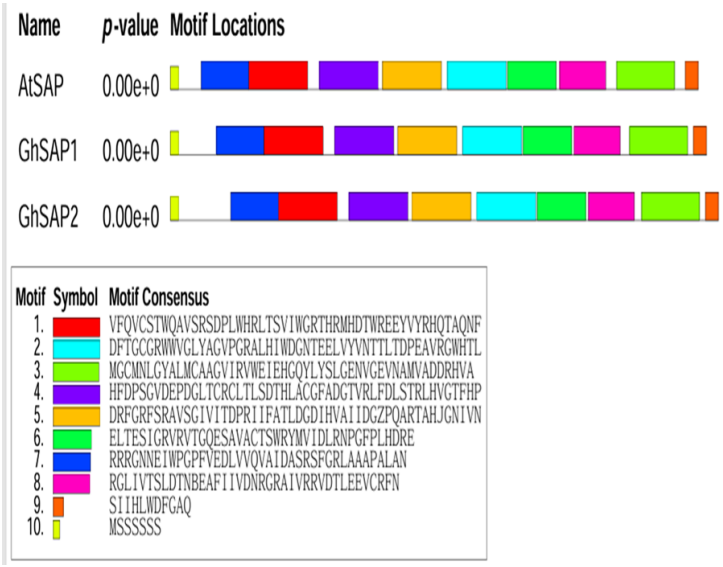

FIGURE 4.5: Consensus motifs of *AtSAP* and *GhSAPs*

4.6 Logos of Identified Motifs

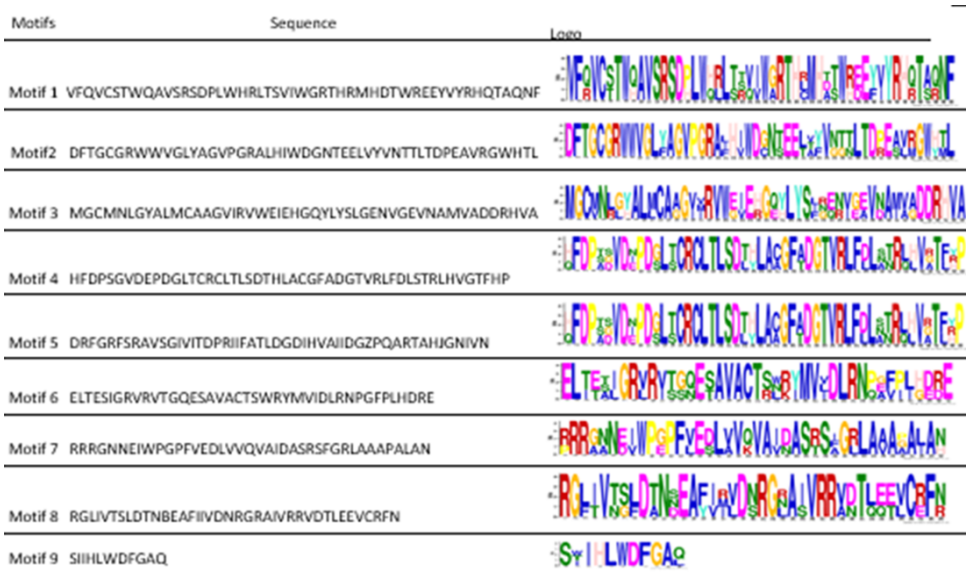

FIGURE 4.6: Logos of Identified Motifs

4.7 Sequence Alignment and Phylogenetic Relationship of *GhSAP* Gene Family

Sequence of different strains of *SAP* genes were taken from different plant species of *Arabidopsis thaliana* and *Gossypium hirsutum*. Sequences were downloaded

and then converted into FASTA file, all sequences were aligned in mega file. Phylogenetic tree is constructed using neighbor-hood joining method. Phylogenetic tree is constructed using CLUSTAL OMEGA (https://www.ebi.ac.uk/clustalo/) and then was redesigned using itol (itol.embl.de/upload.egi).

The evolutionary relationship of *GhSAPs* with different species of *AtSAP* were analyzed by the phylogenetic tree. The *GhSAP1* and *GhSAP2* are diverged in all phylogenetic tree (Fig 4.6). The *SAP* gene family of *Gossypium hirsutum* showed evolutionary relationship with Arabidopsis plant and their structural and functional are predicted to be more similar with *Arabidopsis thaliana* According to the phylogenetic tree.

The phylogenetic tree appears to be divided into four major clades based on visible branching patterns and grouping of sequences.*GhSAP1* (*Gossypium hirsutum SAP1*) is located in Clade III.*GhSAP2* (*Gossypium hirsutum SAP2*) is located in Clade IV. This indicates that these two genes may have diverged significantly and possibly have distinct evolutionary histories or functional specializations.

Closest Relatives of GhSAP1

Gossypium barbadense
Gossypium arboreum
Gossypium mustelinum
These cotton species cluster closely with *GhSAP1* in Clade III, indicating a shared ancestry.

Closest relatives of GhSAP2

Gossypium trilobum
Gossypium klotzschianum
Gossypium lobatum
Gossypium aridum
These are closely grouped with *GhSAP2* in Clade IV, suggesting functional or structural conservation.

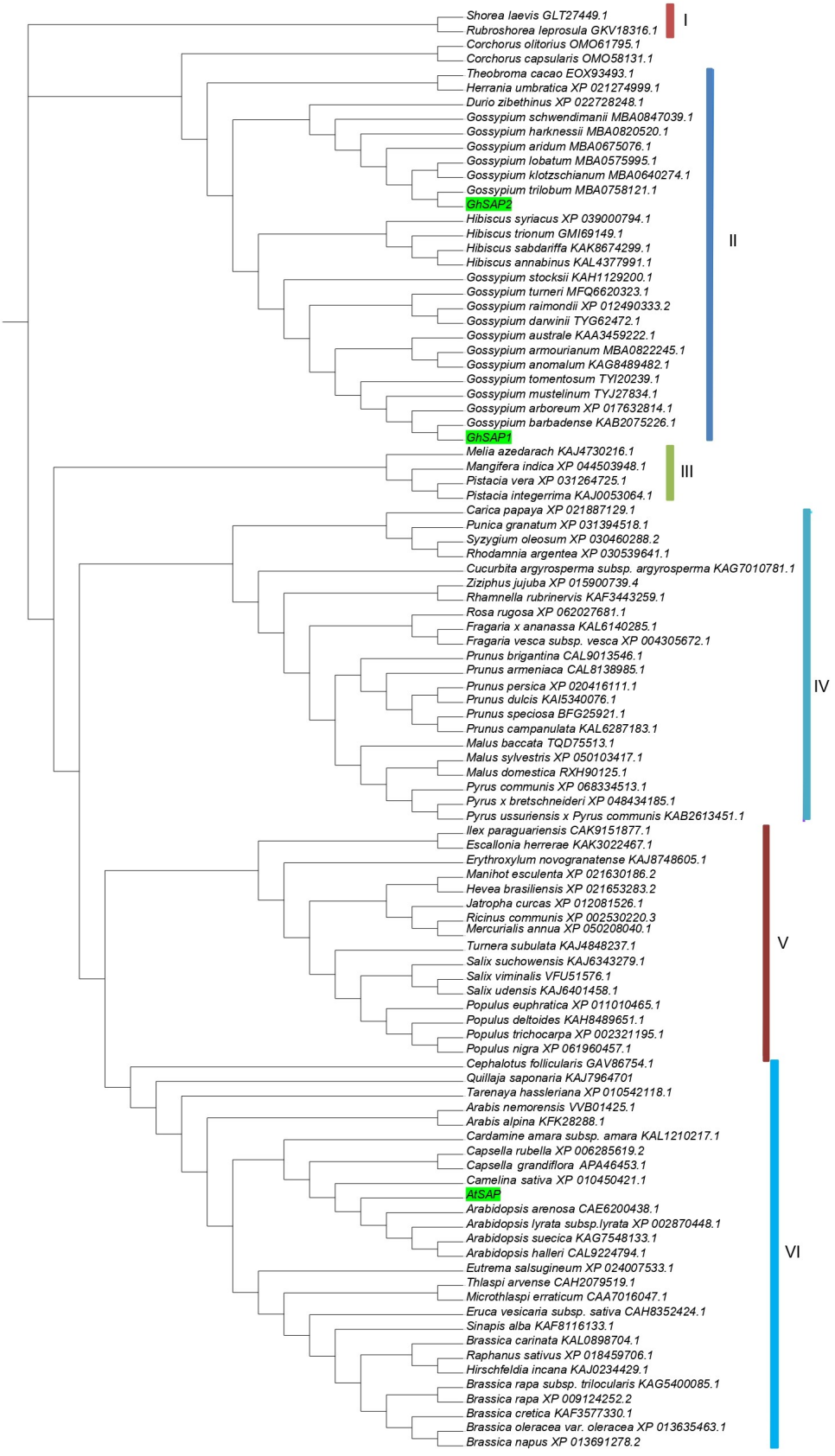

FIGURE 4.7: Phylogenetic analysis of *GhSAPs* through neighbour joining method using MEGA

4.8 Promoter Analysis

One important method for speculating about the function and regulation of genes is the identification of cis-regulatory elements in upstream promoter regions (2000 bp). Three categories of cis-regulatory elements were created in the promoter regions of *AtSAP*, *GhSAP1* and *GhSAP2* in order to classify the identified cis-regulatory elements into three groups: phytohormone (PR), plant growth and development (PGD) and stress (SR)in (Table 4.7).

4.8.1 *AtSAP1*

Shows a broad distribution of regulatory elements, including MRE, CAT-box, circadian, WUN-motif, ARE, LTR, GC-motif, CGTCA-motif, MGACG-motif, GARE-motif, TATC-box, ABRE, ERE, and TCA-.

This suggests that AtSAP1 is highly responsive to diverse signals, such as:

a) Circadian rhythms (circadian motif).

b) Stress responses (ARE, LTR, MRE).

c) Hormonal signals (ABRE for abscisic acid, ERE for ethylene, GARE-motif for gibberellins, CGTCA-motif for jasmonate).

d) Light regulation (TATC-box).

4.8.2 *GhSAP1*

a) Dominated by the Box 4 element (7 copies), which is associated with light responsiveness.

b) Contains circadian elements (2 copies) and ABRE (1 copy), indicating rhythmic and abscisic acid-related regulation.

c) Notably lacks several stress-related motifs (e.g., ARE, LTR) present in At-SAP1.

d) Features a high number of TATC-box (48 copies), strongly implicating light regulation.

4.8.3 *GhSAP2*

a) Shares similarities with GhSAP1 (e.g., Box 4, circadian), but also includes:

b) MRE (2 copies) and ARE (1 copy), suggesting metal and oxidative stress responses.

c) A very high count of GARE-motif (57 copies), pointing to gibberellin hormone signaling.

d) TCA- (2 copies), which may respond to salicylic acid.

Table 4.7: Plant Growth, Stress, and Phytohormone Responsive Motifs in Genes

Gene	**Plant Growth & Development**				**Stress Responsive**				**Phytohormone Responsive**						
	Box 4	**MRE**	**CAT.Box**	**circadian**	**WUN-Motif**	**ARE**	**LTR**	**GC-motif**	**CGTCA-Motif**	**TGACG-Motif**	**GARE-Motif**	**TATC-Box**	**ABRE**	**ERE**	**TCA-**
AtSAP1			1	1		2	1	1	1	1	1	1		1	
GhSAP1	7				1	2						48	1	3	2
GhSAP2	7	2				3		1	1	1	1	57	1	1	2

Chapter 5

Discussion

This study presents a comprehensive genome-wide identification and characterization of the *Sterile Apetala* (*SAP*) gene family in Gossypium hirsutum, employing comparative genomics and cutting-edge bioinformatics tools. By leveraging sequence information from *Arabidopsis thaliana*—a model organism for floral development the study systematically identified and curated putative *SAP* gene homologs in cotton. This approach not only ensured accuracy but also enabled functional inference through evolutionary conservation.

Following a rigorous methodology that involved sequence retrieval, redundancy elimination, and domain validation, two *SAP* genes *GhSAP1* and *GhSAP2* were prioritized for detailed characterization. The selection was based on domain integrity and their phylogenetic relatedness to canonical *SAP* genes in *Arabidopsis*. Their identification is particularly valuable, given the agronomic relevance of cotton and the added complexity introduced by its allotetraploid genome, which often poses challenges in gene identification and functional dissection [140].

Domain architecture analysis revealed that *GhSAP1* and *GhSAP2* harbor F-box and WD40 repeat-like domains, hallmark features of the *SAP* gene family. The presence of these domains supports their involvement in ubiquitin-mediated proteolysis and protein-protein interactions, implicating them in multiple cellular signaling and developmental pathways. The comparative analysis also showed

remarkable similarity in physicochemical properties between *Arabidopsis* and cotton *SAP* proteins, including molecular weight, isoelectric point, and hydrophilic indices. Such conservation reinforces the evolutionary stability and functional importance of *SAP* genes across species [141].

Structural annotation revealed that *GhSAP* genes consist of well-defined exon-intron structures and UTRs, mirroring those in *Arabidopsis*. These features suggest that post-transcriptional regulatory mechanisms, such as alternative splicing or mRNA stability, may be conserved as well. MEME-based motif analysis confirmed the presence of conserved domains essential for *SAP* protein functionality. Notably, the identified motifs reflect regions responsible for protein stability and interaction, underlining their significance in maintaining the structural and functional integrity of *SAP* proteins [142].

Moreover, the acidic nature of these proteins, combined with subcellular localization predictions indicating chloroplast targeting, provides clues about their role in organelle-specific processes. Chloroplasts, being pivotal hubs for photosynthesis and stress sensing, may serve as platforms for *SAP*-mediated regulatory pathways, particularly those involving stress perception, metabolic reprogramming, and floral induction[143].

Protein-protein interaction (PPI) network analysis via STRING revealed that *GhSAP*s are embedded in complex regulatory networks, interacting with both characterized and hypothetical proteins. Some of these interacting partners are known to participate in nitrogen metabolism, hormone signaling, and abiotic stress response pathways, suggesting multifunctional roles for *GhSAP*s. These interactions imply that *SAP* genes could act as molecular integrators, coordinating developmental cues with environmental signals, a trait especially important for crop plants in variable climates.

The phylogenetic tree, constructed using the neighbor-joining method, showed that *GhSAP1* and *GhSAP2* have diverged significantly, clustering into distinct clades with different sets of Gossypium species. *GhSAP1* grouped with *G. barbadense, G. arboreum,* and *G. mustelinum*, while *GhSAP2* clustered with *G.*

trilobum, G. klotzschianum, and *G. lobatum*. This divergence suggests functional specialization among gene paralogs, likely driven by lineage-specific adaptive pressures. The differences observed in gene structure and motif composition between *GhSAP1* and *GhSAP2* further support this hypothesis, pointing toward neo functionalization or sub functionalization after duplication events [144].

Comparative genomic analyses across *Gossypium* species revealed that *SAP* gene family expansion was largely shaped by whole-genome duplication (WGD) and segmental duplication, evolutionary forces well-documented in the cotton lineage. Purifying selection, inferred from synonymous/non-synonymous substitution rates, appears to maintain the structural integrity of essential domains, ensuring the retention of core functions. The conservation across diploid (A and D genome) and tetraploid species underscores the ancestral importance of *SAP* genes in maintaining floral development and reproductive fidelity.

Promoter analysis using PlantCARE databases revealed a rich repertoire of cis-regulatory elements associated with stress responsiveness (e.g., DRE, ARE, HSE), hormonal signaling (e.g., ABRE, TCA, GARE), and floral induction (e.g., CCAAT-box, GT1-motif). These findings imply that *SAP* gene expression is tightly regulated and highly responsive to both endogenous cues and external stimuli. Notably, motifs related to drought, salinity, and thermal stress suggest that *SAP*s could modulate gene expression in response to environmental challenges an especially relevant function given cotton's cultivation in arid and semi-arid regions [145].

This dual role linking developmental regulation with stress adaptability elevates *SAP* genes as potential master regulators in cotton. The capacity of *SAP*s to bridge environmental perception with organ identity decisions is particularly important under climate stress, where reproductive success is tightly coupled with yield stability [146].

Previous functional characterization in *Arabidopsis* has established *SAP*s as key transcriptional regulators of floral patterning, often acting alongside homeotic genes such as *APETALA2*, *AGAMOUS*, and *LEAFY*. The expression of *GhSAP*s in cotton floral tissues, especially during early stages of bud development, suggests

a conserved role in floral meristem determination. Moreover, the co-expression of *GhSAP*s with genes involved in hormonal pathways (like gibberellin and auxin) hints at hormone-mediated cross-talk, potentially influencing boll development, seed set, and fiber initiation [147].

These insights carry important implications for crop improvement. Targeted manipulation of *SAP* genes via CRISPR/Cas genome editing, RNAi, or overexpression may offer novel avenues to improve flowering synchrony, fertility, and stress tolerance, ultimately enhancing boll retention and fiber yield. Such translational strategies are vital for modern breeding programs that aim to develop climate-smart cotton varieties.

This research not only contributes to the understanding of *SAP* gene structure, evolution, and function in *G. hirsutum* but also lays the groundwork for future functional validation and genetic enhancement. The integration of *in silico* characterization with experimental approaches such as transcriptome profiling, promoter-reporter assays, and mutant analysis will be critical to fully unravel the biological significance of *SAP* genes. As the demand for resilient, high-yield cotton cultivars grows in the face of global climate challenges, the insights from this study provide a strategic foundation for precision breeding and sustainable cotton production [148].

Chapter 6

Conclusion

In conclusion, this research provides a comprehensive genome-wide identification and characterization of the *Sterile Apetala* (*SAP*) gene family in *Gossypium hirsutum*, leveraging comparative genomics and advanced bioinformatics tools. By systematically analyzing the structural, evolutionary, and functional attributes of *GhSAP1* and *GhSAP2*, the study uncovers their conserved domains, physicochemical properties, and potential roles in floral development and stress response [129]. Phylogenetic analyses highlight the evolutionary divergence and specialization of these genes within the cotton lineage, while motif and protein-protein interaction studies suggest functional conservation with model plants such as Arabidopsis thaliana. The integration of promoter analysis further elucidates the regulatory complexity governing *SAP* gene expression. Collectively, these findings lay a robust foundation for future functional validation and genetic improvement efforts, offering valuable insights for breeding programs aimed at enhancing cotton yield, resilience, and reproductive success. This work not only advances the understanding of floral regulatory mechanisms in cotton but also sets the stage for translational applications in crop biotechnology and sustainable agriculture.

6.1 Future Recommendations

The technique utilized in this study was found fruitful and very compelling in accomplishing the point of this concentrate hence, this strategy is proposed to be utilized in the investigation of different plants in future. Other quality families might be distinguished utilizing the technique, having modern or business benefits, with low conservative expense and less time utilization. The data uncovered from this study gives a strong hypothetical foundation regarding the matter, nonetheless, the review can be continued in wet-lab for confirmation of the *SAP* gene family of *Gossypium hirsutum* utilizing sub-atomic research facility methods, for example, the Polymerase Chain Response (PCR) [149].

This study clears way for other quality families to be distinguished and modern strategies to be further developed which can, truly, upgrade crop creation, yield and proficient utilization of nitrogen in soil. Also, hereditary designing methods can be further developed which can significantly impact the horticulture area. These improvements might hold promising outcomes for our current circumstance, while refining different modern cycles simultaneously [150].

Bibliography

[1] M. V. Byzova, J. Franken, M. G. M. Aarts, J. de Almeida-Engler, G. Engler, C. Mariani, M. M. Van Lookeren Campagne, and G. C. Angenent, "Arabidopsis STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development," Genes & Development, vol. 13, no. 8, pp. 1002–1014, Apr. 1999, doi: 10.1101/gad.13.8.1002.

[2] J. L. Bowman, D. R. Smyth, and E. M. Meyerowitz, "Genetic interactions among floral homeotic genes of Arabidopsis," Development, vol. 112, pp. 1–20, 1991.

[3] J. L. Bowman, J. Alvarez, D. Weigel, E. M. Meyerowitz, and D. R. Smyth, "Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes," Development, vol. 119, pp. 721–743, 1993.

[4] E. S. Coen and E. M. Meyerowitz, "The war of the whorls: Genetic interactions controlling flower development," Nature, vol. 353, pp. 31–37, 1991.

[5] G. N. Drews, J. L. Bowman, and E. M. Meyerowitz, "Negative regulation of the Arabidopsis gene AGAMOUS by the APETALA2 product," Cell, vol. 65, pp. 991–1002, 1991.

[6] R. C. Elliott, A. S. Betzner, E. Huttner, M. P. Oakes, W. Q. J. Tucker, D. Gerentes, P. Perez, and D. R. Smyth, "AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth," Plant Cell, vol. 8, pp. 155–168, 1996.

[7] J. C. Gaiser, K. Robinson-Beers, and C. S. Gasser, "The Arabidopsis SUPERMAN gene mediates asymmetric growth of the outer integument of ovules," Plant Cell, vol. 7, pp. 333–345, 1995.

[8] A. L. Gashler, S. Swaminathan, and V. P. Sukhatme, "A novel repression module, an extensive activation domain, and a bipartite nuclear localisation signal defined in the immediate-early transcription factor Egr-1," Mol. Cell. Biol., vol. 13, pp. 4556–4571, 1993.

[9] J. Goodrich, P. Puangsomlee, M. Martin, D. Long, E. M. Meyerowitz, and G. Coupland, "A polycomb-group gene regulates homeotic gene expression in Arabidopsis," Nature, vol. 386, pp. 44–51, 1997.

[10] U. Gubler and B. J. Hoffman, "A simple and very efficient method for generating complimentary DNA libraries," Gene, vol. 25, pp. 263–269, 1983.

[11] E. Huala and I. M. Sussex, "LEAFY interacts with floral homeotic genes to regulate Arabidopsis floral development," Plant Cell, vol. 4, pp. 901–913, 1992.

[12] M. Hü lskamp, K. Schneitz, and R. E. Pruitt, "Genetic evidence for a long-range activity that directs pollen tube guidance in Arabidopsis," Plant Cell, vol. 7, pp. 57–64, 1995.

[13] V. F. Irish and I. M. Sussex, "Function of the APETALA1 gene during Arabidopsis floral development," Plant Cell, vol. 2, pp. 741–753, 1990.

[14] U. Janssen-Timmen, P. Lemaire, M.-G. Mattei, O. Revelant, and P. Charnay, "Structure, chromosome mapping and regulation of the mouse zinc-finger gene Krox-24; evidence for a common regulatory pathway for immediate-early serum-response genes," Gene, vol. 80, pp. 325–336, 1989.

[15] K. D. Jofuku, B. G. W. den Boer, M. Van Montagu, and J. K. Okamuro, "Control of Arabidopsis flower and seed development by the homeotic gene APETALA2," Plant Cell, vol. 6, pp. 1211–1225, 1994.

[16] S. A. Kempin, B. Savidge, and M. Yanofsky, "Molecular basis of the cauliflower phenotype in Arabidopsis," Science, vol. 267, pp. 522–525, 1995.

[17] K. M. Klucher, H. Chow, L. Reiser, and R. L. Fischer, "The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2," Plant Cell, vol. 8, pp. 137–153, 1996.

[18] M. K. Komaki, K. Okada, E. Nishino, and Y. Shimura, "Isolation and characterization of novel mutants of Arabidopsis thaliana defective in flower development," Development, vol. 104, pp. 195–203, 1988.

[19] L. Kunst, J. E. Klenz, J. Martinez-Zapater, and G. W. Haughn, "AP2 gene determines the identity of perianth organs in flowers of Arabidopsis thaliana," Plant Cell, vol. 1, pp. 1195–1208, 1989.

[20] J. D. Lang, S. Ray, and A. Ray, "sin1, a mutation affecting female fertility in Arabidopsis, interacts with mod1, its recessive modifier," Genetics, vol. 137, pp. 1101–1110, 1994.

[21] K. M. Leon-Kloosterziel, C. J. Keijzer, and M. Koornneef, "A seed shape mutant of Arabidopsis that is affected in integument development," Plant Cell, vol. 6, pp. 385–392, 1994.

[22] Z. Liu and E. M. Meyerowitz, "LEUNIG regulates AGAMOUS expression in Arabidopsis flowers," Development, vol. 121, pp. 975–991, 1995.

[23] M. A. Mandel and M. F. Yanofsky, "A gene triggering flower formation in Arabidopsis," Nature, vol. 377, pp. 522–524, 1995.

[24] M. A. Mandel, C. Gustafson-Brown, B. Savidge, and M. F. Yanofsky, "Molecular characterization of the Arabidopsis floral homeotic gene APETALA1," Nature, vol. 360, pp. 273–277, 1992.

[25] P. J. Mitchell and R. Tjian, "Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins," Science, vol. 245, pp. 371–378, 1989.

[26] Y. Mizukami and H. Ma, "Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity," Cell, vol. 71, pp. 119–131, 1992.

[27] S. Aslam, S. H. Khan, A. Ahmed, and A. M. Dandekar, "The Tale of Cotton Plant: From Wild Type to Domestication, Leading to Its Improvement by Genetic Transformation," American Journal of Molecular Biology, vol. 10, no. 2, pp. 91-111, Feb. 2020, doi: 10.4236/ajmb.2020.102008.

[28] T. Banuri, "Pakistan: Environmental Impact of Cotton Production and Trade," Int. Institute of Sustainable Development, Canada, 1998.

[29] C. B. Cororaton and D. Orden, "Pakistan's Cotton and Textile Economy, Intersectoral Linkages and Effects on Rural and Urban Poverty," International Food Policy Research Institute, USA, 2008.

[30] Economic Survey of Pakistan, Ministry of Food, Agriculture and Livestock, Economic Advisor Wing, Islamabad, Pakistan, 2017.

[31] Office of the Gene Technology Regulator, "The Biology of Gossypium hirsutum L. and Gossypium barbadense L. (Cotton)," Australian Government, Department of Health and Ageing, 2008. . Available: http://www.ogtr.gov.au

[32] J. Quinn and D. Kelly, "Australian Cotton Production Manual," Crop Growth Stages, vol. 7, pp. 44–47, 2011.

[33] B. Robertson, C. Bednarz, and C. Burmester, "Growth and Development—First 60 Days," Newsletter of the Cotton Physiology Education Program, National Cotton Council, vol. 13, pp. 1–5, 2007.

[34] D. M. Oosterhuis and J. L. Snider, "High Temperature Stress on Floral Development and Yield of Cotton," in High Temperature Stress on Floral Development and Yield of Cotton, National Cotton Council of America, 2010. . Available: https://www.cotton.org/foundation/upload/Stress-Physiology-in-Cotton_Chapter1.pdf

[35] J. F. Wendel et al., "Evolution and Natural History of the Cotton Genus," in Genetics and Genomics of Cotton, A. H. Paterson, Ed., Springer, 2009, pp. 3–22.

[36] J. F. Wendel and C. E. Grover, "Taxonomy and Evolution of the Cotton Genus, Gossypium," in Cotton, 2nd ed., American Society of Agronomy, 2015, pp. 1–20.

[37] Z. J. Chen et al., "Towards Sequencing Cotton (Gossypium) Genomes," Plant Physiology, vol. 145, pp. 1303–1310, 2007.

[38] D. R. Duncan, "Cotton Transformation," in Cotton Biotechnology in Agriculture and Forestry, vol. 65, Springer, 2010, pp. 1–20.

[39] B. Schoen, The Fragile Fabric of Union: Cotton, Federal Politics and the Global Origins of the Civil War, Johns Hopkins University Press, 2009.

[40] I. S. Chaudhry, M. B. Khan, and M. Anwar, "Factors Affecting Cotton Production in Pakistan: Empirical Evidence from Multan District," Journal of Quality and Technology Management, vol. V, no. II, pp. 91-100, Dec. 2009.

[41] E. Brite and J. M. Marston, "Environmental Change, Agricultural Innovation, and the Spread of Cotton Agriculture in the Old World," Journal of Anthropology and Archaeology, vol. 32, pp. 39–53, 2013.

[42] C. Moulherat et al., "First Evidence of Cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead," Journal of Archaeological Science, vol. 29, pp. 1393–1401, 2002.

[43] G. Hustvedt and P. C. Crew, "Textile Technology," The Journal of Cotton Science, vol. 9, pp. 47–55, 2005.

[44] J. M. Vreeland Jr., "The Revival of Colored Cotton," Scientific American, vol. 280, p. 112, 1999.

[45] W. Malik, A. A. Khan, and B. Sadia, "In Situ Characterization of Coloured Cotton Genotypes," Australian Journal of Crop Science, vol. 7, pp. 299–304, 2013.

[46] A. I. Khan, Y. B. Fu, and I. A. Khan, "Genetic Diversity of Pakistani Cotton Cultivars as Revealed by Simple Sequence Repeat Markers," Communications in Biometry and Crop Science, vol. 4, pp. 21–30, 2009.

[47] W. Z. Guo, K. Wang, and T. Z. Zhang, "A and D Genome Evolution in Gossypium Revealed Using SSR Molecular Markers," Journal of Genetics and Genomics, vol. 30, pp. 183–188, 2003.

[48] X. Zhao et al., "Macromolecular Organization and Genetic Mapping of a Rapidly Evolving Chromosome-Specific Tandem Repeat Family (B77) in Cotton (Gossypium)," Plant Molecular Biology, vol. 38, pp. 1031–1042, 1998.

[49] J. T. Page et al., "Insights into the Evolution of Cotton Diploids and Polyploids from Whole-Genome Re-Sequencing," G3, vol. 3, pp. 1809–1818, 2013.

[50] K. Wang et al., "The Draft Genome of a Diploid Cotton Gossypium raimondii," Nature Genetics, vol. 44, pp. 1098–1103, 2012.

[51] W. Guo et al., "A Preliminary Analysis of Genome Structure and Composition in Gossypium hirsutum," BMC Genomics, vol. 9, p. 314, 2008.

[52] A. H. Paterson et al., "Repeated Polyploidization of Gossypium Genomes and the Evolution of Spinnable Cotton Fibres," Nature, vol. 492, pp. 423–427, 2012.

[53] Z. Wang et al., "A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of Gossypium raimondii L.," G3, vol. 3, pp. 1759–1767, 2013.

[54] A. Blenda et al., "A High Density Consensus Genetic Map of Tetraploid Cotton That Integrates Multiple Component Maps through Molecular Marker Redundancy Check," PLoS ONE, vol. 7, p. e45739, 2012.

[55] W. Xie et al., "Parent-Independent Genotyping for Constructing an Ultrahigh-Density Linkage Map Based on Population Sequencing," Proc. Natl. Acad. Sci. USA, vol. 107, pp. 10578–10583, 2010.

[56] Z. Sun et al., "An Ultradense Genetic Recombination Map for Brassica napus, Consisting of 13,551 SRAP Markers," Theoretical and Applied Genetics, vol. 114, pp. 1305–1317, 2007.

[57] F. Li et al., "Genome Sequence of the Cultivated Cotton Gossypium arboreum," Nature Genetics, vol. 46, pp. 567–572, 2014.

[58] F. Li et al., "Genome Sequence of Cultivated Upland Cotton (Gossypium hirsutum TM-1) Provides Insights into Genome Evolution," Nature Biotechnology, vol. 33, pp. 524–530, 2015.

[59] B. Hendrix and J. M. Stewart, "Estimation of the Nuclear DNA Content of Gossypium Species," Annals of Botany, vol. 95, pp. 789–797, 2005.

[60] T. Zhang et al., "Sequencing of Allotetraploid Cotton (Gossypium hirsutum L. acc. TM-1) Provides a Resource for Fiber Improvement," Nature Biotechnology, vol. 33, pp. 531–537, 2015.

[61] X. Liu et al., "Gossypium barbadense Genome Sequence Provides Insight into the Evolution of Extra-Long Staple Fiber and Specialized Metabolites," Scientific Reports, vol. 5, p. 14139, 2015.

[62] B.-H. Wang et al., "QTL Mapping for Plant Architecture Traits in Upland Cotton Using RILs and SSR Markers," Acta Genetica Sinica, vol. 33, pp. 161–170, 2006.

[63] I. Y. Abdurakhmonov et al., "Molecular Diversity and Association Mapping of Fiber Quality Traits in Exotic G. hirsutum L. Germplasm," Genomics, vol. 92, pp. 478–487, 2008.

[64] C. H. C. de Magalhaes Bertini et al., "Characterization and Genetic Diversity Analysis of Cotton Cultivars Using Microsatellites," Genetics and Molecular Biology, vol. 29, pp. 321–329, 2006.

[65] G. Chen and X.-M. Du, "Genetic Diversity of Source Germplasm of Upland Cotton in China as Determined by SSR Marker Analysis," Acta Genetica Sinica, vol. 33, pp. 733–745, 2006.

[66] P. Tyagi et al., "Genetic Diversity and Population Structure in the US Upland Cotton (Gossypium hirsutum L.)," Theoretical and Applied Genetics, vol. 127, pp. 283–295, 2014.

[67] B. Ehsan et al., "Assessment of Genomic Diversity of Cotton (Gossypium hirsutum) Genotypes Using Simple Sequence Repeats Markers through Genetic Analysis Software," International Journal of Agriculture and Biology, vol. 15, pp. 968–972, 2013.

[68] W. Nazeer et al., "Introgression of Genes for Cotton Leaf Curl Virus Resistance and Increased Fiber Strength from Gossypium stocksii into Upland Cotton (G. hirsutum)," Genetics and Molecular Research, vol. 13, pp. 1133–1143, 2014.

[69] R. Mann, "Bemisia tabaci Interaction with Cotton Leaf Curl Virus," in The Whitefly, Bemisia tabaci (Homoptera. Aleyrodidae) Interaction with Geminivirus-Infected Host Plants, W. M. O. Thompson, Ed., Springer, 2011, pp. 69–88.

[70] D. Singh et al., "Yield and Fibre Quality Associated with Cotton Leaf Curl Disease of Bi-Cotton in Punjab," Journal of Environmental Biology, vol. 34, pp. 113–116, 2013.

[71] S. Ahmad et al., "CRSM-38, a New High Yielding Coupled with CLCuV Tolerance Cotton (Gossypium hirsutum L.) Variety," African Journal of Biotechnology, vol. 11, pp. 4368–4677, 2012.

[72] J. H. Cai et al., "Cotton Leaf Curl Multan Virus Newly Reported to Be Associated with Cotton Leaf Curl Disease in China," Plant Pathology, vol. 59, pp. 794–795, 2010.

[73] M. N. Sattar et al., "Cotton Leaf Curl Disease—An Emerging Threat to Cotton Production Worldwide," Journal of General Virology, vol. 94, pp. 695–710, 2013.

[74] M. T. Azhar, S. Akhtar, and S. Mansoor, "Cotton Leaf Curl Multan Betaastellite Strains Cloned from Gossypium barbadense Further Supports Selection Due to Host Resistance," Virus Genes, vol. 45, pp. 402–405, 2012.

[75] C. Kirda et al., "Prospects of Partial Root Zone Irrigation for Increasing Irrigation Water Use Efficiency of Major Crops in the Mediterranean Region," Annals of Applied Biology, vol. 150, pp. 281–291, 2007.

[76] N. C. Ellstrand, H. C. Prentice, and J. F. Hancock, "Gene Flow and Introgression from Domesticated Plants into Their Wild Relatives," Annual Review of Ecology and Systematics, vol. 30, pp. 539–563, 1999.

[77] J. J. Bull and H. A. Wichman, "Applied Evolution," Annual Review of Ecology and Systematics, vol. 32, pp. 183–217, 2001.

[78] R. W. Allard, "History of Plant Population Genetics," Annual Review of Genetics, vol. 33, pp. 1–27, 1999.

[79] J. Zhang et al., "Genetic Improvement of New Mexico Acala Cotton Germplasm and Their Genetic Diversity," Crop Science, vol. 45, pp. 2363–2373, 2005.

[80] A. Chowdhary, S. Mishra, S. Mehrotra, S. K. Upadhyay, D. Bagal, and V. Srivastava, "Plant transcription factors: an overview of their role in plant life," Book Chapter, Department of Botany, School of Life Sciences, Central University of Jammu, Samba, Jammu and Kashmir (UT), India, 2023. doi: 10.1016/b978-0-323-90613-5.00003-01.

[81] M. Agarwal et al., "A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance," J. Biol. Chem., vol. 281, pp. 37636–37645, 2006.

[82] M. Akhtar et al., "DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants," J. Genet., vol. 91, pp. 385–395, 2012.

[83] M. A. Ali et al., "Transcription factors WRKY11 and WRKY17 are involved in abiotic stress responses in Arabidopsis," J. Plant Physiol., vol. 226, pp. 12–21, 2018.

[84] J. P. Alvarez et al., "Active suppression of a leaf meristem orchestrates determinate leaf growth," Elife, vol. 5, p. e15023, 2016.

[85] E. H. Baillo et al., "Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement," Genes, vol. 10, no. 10, p. 771, 2019.

[86] E. Baldoni, A. Genga, and E. Cominelli, "Plant MYB transcription factors: their role in drought response mechanisms," Int. J. Mol. Sci., vol. 16, no. 7, pp. 15811–15851, 2015.

[87] M. Bar and N. Ori, "Leaf development and morphogenesis," Development, vol. 141, no. 22, pp. 4219–4230, 2014.

[88] C. Bernhardt et al., "The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root," Development, vol. 130, pp. 6431–6439, 2003.

[89] R. Bhatia et al., "Modified expression of ZmMYB167 in Brachypodium distachyon and Zea mays leads to increased cell wall lignin and phenolic content," Sci. Rep., vol. 9, p. 1, 2019.

[90] R. Borner et al., "A MADS domain gene involved in the transition to flowering in Arabidopsis," Plant J., vol. 24, no. 5, pp. 591–599, 2000.

[91] J. Bou-Torrent et al., "ATHB4 and HAT3, two class II HD-ZIP transcription factors, control leaf development in Arabidopsis," Plant Signal. Behav., vol. 7, pp. 1382–1387, 2012.

[92] R. Cai et al., "The maize WRKY transcription factor ZmWRKY17 negatively regulates salt stress tolerance in transgenic Arabidopsis plants," Planta, vol. 246, pp. 1215–1231, 2017.

[93] H. Cao et al., "Tomato transcriptional repressor MYB70 directly regulates ethylene-dependent fruit ripening," Plant J., vol. 104, pp. 1568–1581, 2020.

[94] J. L. Celenza et al., "The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis," Plant Physiol., vol. 137, pp. 253–262, 2005.

[95] H. C. Chen et al., "The transcription factor OsbHLH035 mediates seed germination and enables seedling recovery from salt stress through ABA-dependent and ABA-independent pathways, respectively," Rice, vol. 11, p. 1, 2018.

[96] Y. H. Chen et al., "Identification of MYB transcription factor genes and their expression during abiotic stresses in maize," Biol. Plant., vol. 62, pp. 222–230, 2018.

[97] R. S. Chiu et al., "The role of the Arabidopsis FUSCA3 transcription factor during inhibition of seed germination at high temperature," BMC Plant Biol., vol. 12, p. 1, 2012.

[98] Y. Dai et al., "Isolation and characterization of a novel cDNA encoding methyl jasmonate-responsive transcription factor TcAP2 from Taxus cuspidata," Biotechnol. Lett., vol. 31, pp. 1801–1809, 2009.

[99] C. Deng et al., "The rice transcription factors OsICE confer enhanced cold tolerance in transgenic Arabidopsis," Plant Signal. Behav., vol. 12, no. 5, p. e1316442, 2017.

[100] C. Deng et al., "ABA-responsive transcription factor bZIP1 is involved in modulating biosynthesis of phenolic acids and tanshinones in Salvia miltiorrhiza," J. Exp. Bot., vol. 71, pp. 5948–5962, 2020.

[101] Y. Duran-Medina et al., "The AP2/ERF transcription factor DRNL modulates gynoecium development and affects its response to cytokinin," Front. Plant Sci., vol. 8, p. 1841, 2017.

[102] M. A. El-Esawi and A. A. Alayafi, "Overexpression of StDREB2 transcription factor enhances drought stress tolerance in cotton (Gossypium barbadense L.)," Genes, vol. 10, p. 142, 2019.

[103] M. A. El-Esawi et al., "Overexpression of AtWRKY30 transcription factor enhances heat and drought stress tolerance in wheat (Triticum aestivum L.)," Genes, vol. 10, no. 2, p. 163, 2019.

[104] A. El-Kereamy et al., "The rice R2R3-MYB transcription factor OsMYB55 is involved in the tolerance to high temperature and modulates amino acid metabolism," PLoS One, vol. 7, p. e52030, 2012.

[105] R. V. Espley et al., "Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10," Plant J., vol. 49, pp. 414–427, 2007.

[106] C. Fan et al., "Genome-wide analysis of Eucalyptus grandis WRKY genes family and their expression profiling in response to hormone and abiotic stress treatment," Gene, vol. 678, pp. 38–48, 2018.

[107] Y. Fang et al., "A stress-responsive NAC transcription factor SNAC3 confers heat and drought tolerance through modulation of reactive oxygen species in rice," J. Exp. Bot., vol. 66, pp. 6803–6817, 2015.

[108] R. Gao et al., "SPL13 regulates shoot branching and flowering time in Medicago sativa," Plant Mol. Biol., vol. 96, pp. 119–133, 2018.

[109] H. Gao et al., "Overexpression of a WRKY transcription factor TaWRKY2 enhances drought stress tolerance in transgenic wheat," Front. Plant Sci., vol. 9, p. 997, 2018.

[110] F. Giorno et al., "Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers," J. Exp. Bot., vol. 61, pp. 453–462, 2010.

[111] R. Borner, G. Kampmann, J. Chandler, R. Gleissner, E. Wisman, K. Apel, and G. Coupland, "A MADS domain gene involved in the transition to flowering in Arabidopsis," The Plant Journal, vol. 24, no. 5, pp. 591–599, 2000.

[112] M. A. Mandel, C. Gustafson-Brown, B. Savidge, and M. F. Yanofsky, "Molecular characterization of the Arabidopsis floral homeotic gene APETALA1," Nature, vol. 360, pp. 273–277, 1992.

[113] D. Weigel, J. Alvarez, D. R. Smyth, M. F. Yanofsky, and E. M. Meyerowitz, "LEAFY controls floral meristem identity in Arabidopsis," Cell, vol. 69, pp. 843–859, 1992.

[114] G. B. Seymour, L. Ostergaard, N. H. Chapman, S. Knapp, and C. Martin, "Fruit development and ripening," Annual Review of Plant Biology, vol. 64, pp. 219–241, 2013.

[115] J. Paz-Ares, D. Ghosal, U. Wienand, P. A. Peterson, and H. Saedler, "The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators," The EMBO Journal, vol. 6, pp. 3553–3558, 1987.

[116] R. V. Espley, R. P. Hellens, J. Putterill, D. E. Stevenson, S. Kutty-Amma, and A. C. Allan, "Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10," The Plant Journal, vol. 49, pp. 414–427, 2007.

[117] H. Cao, J. Chen, M. Yue, C. Xu, W. Jian, Y. Liu, and J. Chen, "Tomato transcriptional repressor MYB70 directly regulates ethylene-dependent fruit ripening," The Plant Journal, vol. 104, pp. 1568–1580, 2020.

[118] G. Khaksar, W. Sangchay, P. Pinsorn, L. Sangpong, and S. Sirikantaramas, "Genome-wide analysis of the Dof gene family in durian reveals fruit ripening-associated and cultivar-dependent Dof transcription factors," Scientific Reports, vol. 9, pp. 1–13, 2019.

[119] X. Xie, Y. Ma, P. Bi, W. Wei, J. Liu, Y. Hu, and Y. Zhang, "Transcription factor FvTCP9 promotes strawberry fruit ripening by regulating the biosynthesis of abscisic acid and anthocyanins," Plant Physiology and Biochemistry, vol. 146, pp. 374–383, 2020.

[120] N. Ma, H. Feng, X. Meng, D. Li, D. Yang, C. Wu, and Q. Meng, "Overexpression of tomato SINAC1 transcription factor alters fruit pigmentation and softening," BMC Plant Biology, vol. 14, pp. 1–14, 2014.

[121] M. Bar and N. Ori, "Leaf development and morphogenesis," Development, vol. 141, no. 22, pp. 4219–4230, 2014.

[122] K. Sarvepalli and U. Nath, "Hyper-activation of the TCP4 transcription factor in Arabidopsis thaliana accelerates multiple aspects of plant maturation," The Plant Journal, vol. 67, pp. 595–607, 2011.

[123] J. Bou-Torrent, M. Salla-Martret, R. Brandt, T. Musielak, J. C. Palauqui, and J. F. Martinez-Garcia, "ATHB4 and HAT3, two class II HD-ZIP transcription factors, control leaf development in Arabidopsis," Plant Signaling & Behavior, vol. 7, pp. 1382–1387, 2012.

[124] M. Omidbakhshfard, S. Proost, U. Fujikura, and B. Mueller-Roeber, "Growth-regulating factors (GRFs): a small transcription factor family with important functions in plant biology," Molecular Plant, vol. 8, pp. 998–1010, 2015.

[125] J. P. Alvarez, C. Furumizu, I. Efroni, Y. Eshed, and J. L. Bowman, "Active suppression of a leaf meristem orchestrates determinate leaf growth," Elife, vol. 5, p. e15023, 2016.

[126] A. Hasson, A. Plessis, T. Blein, B. Adroher, S. Grigg, M. Tsiantis, and P. Laufs, "Evolution and diverse roles of the CUP-SHAPED COTYLEDON genes in Arabidopsis leaf development," The Plant Cell, vol. 23, pp. 54–68, 2011.

[127] Q. Xie, G. Frugis, D. Colgan, and N. H. Chua, "Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development," Genes & Development, vol. 14, pp. 3024–3036, 2000.

[128] H. Zhang and B. G. Forde, "An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture," Science, vol. 279, pp. 407–409, 1998.

[129] C. Bernhardt, M. M. Lee, A. Gonzalez, F. Zhang, A. Lloyd, and J. Schiefelbein, "The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root," Development, vol. 130, pp. 6431–6439, 2003.

[130] J. D. Masucci, W. G. Rerie, D. R. Foreman, M. Zhang, M. E. Galway, and M. D. Marks, "The homeobox gene GLABRA2 is required for position-dependent cell differentiation in the root epidermis of Arabidopsis thaliana," Development, vol. 122, pp. 1253–1260, 1996.

[131] T. Wada, T. Tachibana, Y. Shimura, and K. Okada, "Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC," Science, vol. 277, pp. 1113–1116, 1997.

[132] V. Kirik, M. Simon, M. Huelskamp, and J. Schiefelbein, "The ENHANCER OF TRY AND CPC1 gene acts redundantly with TRIPTYCHON and CAPRICE in trichome and root hair cell patterning in Arabidopsis," Developmental Biology, vol. 268, pp. 506–513, 2004.

[133] S. Schellmann, A. Schnittger, V. Kirik, T. Wada, K. Okada, A. Beermann, and M. Hulskamp, "TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis," The EMBO Journal, vol. 21, pp. 5036–5046, 2002.

[134] D. Yu, T. Fang, Z. Tu, X. Guo, and W. Zhang, "CottonGen: A genomics, genetics and breeding database for cotton research," The Plant and Animal Genome Conference XXVI, San Diego, CA, USA, Jan. 2018. . Available: https://www.cottongen.org

[135] M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, and B. Morgenstern, "AUGUSTUS: Ab initio prediction of alternative transcripts," Nucleic Acids Research, vol. 34, no. suppl_2, pp. W435–W439, Jul. 2006.

[136] M. Hoff, A. Lomsadze, M. Borodovsky, and M. Stanke, "BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS," Bioinformatics, vol. 32, no. 5, pp. 767–769, Mar. 2016.

[137] E. M. Emms and S. Kelly, "OrthoFinder: Phylogenetic orthology inference for comparative genomics," Genome Biology, vol. 20, no. 1, p. 238, 2019.

[138] R. D. Finn, J. Clements, and S. R. Eddy, "HMMER web server: Interactive sequence similarity searching," Nucleic Acids Research, vol. 39, no. suppl_2, pp. W29–W37, Jul. 2011.

[139] F. J. Theis and M. Tsoucas, "Integrating single-cell multi-omics data using deep learning," Nature Reviews Genetics, vol. 23, pp. 36–54, 2022.

[140] W. Hu, Y. Chen, H. Zhang, X. Liu, Y. Yu, and X. Gao, "Genome-wide identification and expression analysis of the SAP gene family in upland cotton (Gossypium hirsutum)," International Journal of Molecular Sciences, vol. 21, no. 1, pp. 1–17, Jan. 2020, doi: 10.3390/ijms21010201.

[141] A. Ahmad, S. Zafar, M. Hussain, S. Khan, and F. Anwar, "Structural and functional analysis of SAP gene family in cotton reveals their roles in stress response," Plant Physiology and Biochemistry, vol. 167, pp. 112–122, Nov. 2021.

[142] T. Bailey et al., "MEME SUITE: Tools for motif discovery and searching," Nucleic Acids Research, vol. 37, no. suppl_2, pp. W202–W208, Jul. 2009, doi: 10.1093/nar/gkp335.

[143] Y. Wang, L. Zhang, C. Chen, and Y. Zhang, "Subcellular localization and expression analysis of stress-related genes in G. hirsutum," BMC Plant Biology, vol. 18, no. 1, p. 232, 2018.

[144] X. Zhang, J. Guo, H. Wang, and L. Gao, "Evolutionary expansion and functional divergence of SAP gene family in Gossypium species," Frontiers in Genetics, vol. 12, pp. 1–14, 2021.

[145] L. Lescot et al., "PlantCARE: A database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences," Nucleic Acids Research, vol. 30, no. 1, pp. 325–327, 2002.

[146] M. Yamaguchi and S. Abe, "Roles of SAP transcription factors in flower development and stress responses," Plant Cell Reports, vol. 39, pp. 1441–1453, 2020.

[147] M. Ferrándiz, "Regulation of floral meristem identity and determinacy in Arabidopsis," Annual Review of Plant Biology, vol. 68, pp. 289–315, 2017.

[148] R. Raza, N. Iqbal, and M. Nawaz, "CRISPR/Cas9-mediated gene editing in cotton: Innovations and future prospects," Plant Biotechnology Journal, vol. 20, pp. 1205–1218, 2022.

[149] K. Edwards et al., "PCR-based marker systems for analyzing cotton gene expression," Journal of Cotton Science, vol. 23, no. 3, pp. 280–289, 2019.

[150] S. K. Jha and D. Singh, "Advances in plant genetic engineering: Applications in agriculture and biotechnology," Journal of Plant Biochemistry and Biotechnology, vol. 31, pp. 200–215, 2022.