Original Article

J Genet Appl Biotechnol, 2026: *e2026040*
DOI: https://doi.org/10.66432/xe71vp03
Received: 29 Apr 2026; Revised: 10 Jun 2026; Accepted: 23 Jun 2026; Published online: 28 Aug 2026

Identification of candidate genes and transcripts through *in silico* analysis of DNA markers

Feruza U. Rafieva[1], Barno B. Oripova[1,2], Abduqahhor Kh. Toshpulatov[1,2], Sevara K. Arslanova[1], Madina D. Kholova[1], Dilrabo K. Ernazarova[1,3], Fakhriddin N. Kushanov[1,3]

[1]Institute of Genetics and Plant Experimental Biology, Yukori-Yuz, Qibray District, Tashkent Region, 111208, Uzbekistan. Tel: (71) 264-23-90.
[2]Tashkent State Medical University, 2 Farobiy Street, Tashkent, 100109, Uzbekistan. Tel: (95) 475-45-01.
[3]National University of Uzbekistan, Universitet Street, Tashkent, 100174, Uzbekistan. Tel: (71) 246-02-24
[*]Corresponding author: Rafieva U. Feruza; feruzarafiyeva25@gmail.com

Abstract

This article presents the results of determining candidate genes using *in silico* PCR using certain microsatellite DNA markers genetically linked to the genome of *G. hirsutum* L. and economically valuable traits. In this study, a virtual analysis of the region (100 kbp) around polymorphic microsatellite markers was performed using the NCBI database, UniPro Ugene, AUGUSTUS, and BLAST software. As a result, 29 valuable candidate genes and proteins were identified during the blast analysis of 10 SSR DNA markers related to fiber quality and salt resistance. The functions of the identified candidate genes and proteins were highlighted based on the interpretation of research results by foreign scientists. These results serve as an important basis for increasing the efficiency of marker-based selection (MAS) in cotton breeding and creating stress-resistant varieties.

Keywords: Cotton, SSR Markers, *In silico* PCR, Candidate Genes, Fiber Quality, Abiotic Stress

How to cite this article: Rafieva FU, Oripova BB, Toshpulatov AK, Arslanova SK, Kholova MD, Ernazarova DK, Kushanov FN. Identification of candidate genes and transcripts through in silico analysis of DNA markers. J Genet Appl Biotechnol. 2026; e2026040. https://doi.org/10.66432/xe71vp03

Introduction

The identification of genes within plant genomes, particularly the cotton genome, and the determination of their functions based on gene or protein sequence analysis using bioinformatics programs and virtual analyses, holds significant scientific and practical importance. Currently, the genomes of 34 diploid (2n=2x=26) and tetraploid (2n=4x=52) cotton species belonging to the genus *Gossypium* L. have been sequenced and uploaded to the NCBI (National Center for Biotechnology Information) genome database (https://www.ncbi.nlm.nih.gov/). The sequential sequencing of the genomes of species from the *Gossypium* L. genus has created the opportunity to identify orthologs of genes responsible for a number of valuable traits in cotton through bioinformatic analysis (1,2). In particular, the sequencing of the medium-staple *G. hirsutum* L. genome has made it possible to identify genes related to valuable agronomic traits in marked genomic regions using bioinformatics methods and genetic databases (3). Although the genomes of 34 diploid and tetraploid species of the *Gossypium* L. genus have been fully sequenced, the mechanisms of fiber quality, resistance to various stress factors and diseases, as well as the genes responsible for these traits, have not been completely identified. Therefore, research in this area is ongoing (4). From this perspective, in our research, we aimed to identify candidate genes and proteins in the cotton genome associated with microsatellite markers that have demonstrated high polymorphism in molecular analyses.

Materials and Methods

This research utilized virtual PCR in the UniPro Ugene software version 1.21 (5), *In silico* PCR analysis, the AUGUSTUS version 3.5.0 web application (6), and BLAST version 2.16.0 (7) search programs. Specifically, virtual PCR was performed in the UniPro Ugene software using the genome of the cotton species *G. hirsutum* L. and SSR DNA markers that are genetically linked to valuable traits. As a result, the hybridization or lack thereof of these markers at specific positions on different chromosomes of the cotton genome was confirmed based on the synthesized PCR amplicons of varying base pair lengths. Virtual analyses were conducted with DNA markers genetically linked to fiber quality indicators and traits associated with resistance to various stress factors and diseases. *In silico* PCR analysis

was used to determine the chromosomal positions of these markers. Using the nucleotide sequences from these designated genomic regions, a probable gene, its transcript variants, and associated proteins were identified based on the Arabidopsis thaliana model plant via the AUGUSTUS web application. Subsequently, important genes and proteins specific to the cotton genome were identified by searching for candidate genes and proteins using the BLAST program. The role of these candidate genes and proteins in various physiological and biochemical processes related to the growth and development of plants, including cotton, was analyzed through a literature review.

In this research, the genome of the *G. hirsutum* L. species, which was uploaded to the NCBI (National Center for Biotechnology Information) genome database, was downloaded, and polymorphic SSR DNA markers genetically linked to valuable traits were utilized. The genome of the species *G. hirsutum* L., embedded in the genomic database, was downloaded, and polymorphic SSR DNA markers genetically linked to valuable traits were used. The selection of SSR markers was based on published literature reporting their association with economically important cotton traits. Specifically, the markers BNL3436, BNL1167, BNL3601, BNL2650, BNL2569, CGR5883, CGR5732, CGR5866, DPL0322, BNL1414, BNL1694, and CIR0246 were chosen because they are located within QTL regions associated with fiber quality traits, resistance to Fusarium wilt, and tolerance to salinity stress, and have been reported to exhibit a high level of polymorphism. These markers provide an effective tool for assessing genetic diversity among genotypes and for evaluating allelic variation linked to target agronomic traits.

Results

A virtual PCR was performed in the Unipro UGENE software using the genome of the cotton species *G. hirsutum* L. and SSR DNA markers genetically linked to valuable traits. Analysis of the results confirms that these markers hybridized at specific positions on different chromosomes of the cotton genome, which is evidenced by the synthesized PCR amplicons of varying base pair lengths (Figure 1).

In addition, virtual analyses were conducted not only with DNA markers linked to fiber quality indicators but also with those genetically linked to traits associated with resistance to various stress factors and diseases. The *in silico* PCR analysis determined the chromosomal positions of these markers (Table 1). Here, FASTA-PCR analysis of SSR DNA markers involved in the regulation of fiber quality and other valuable agronomic traits in the cotton plant, as well as markers based on the cotton genome, shows that the microsatellite markers are located on the corresponding homologous chromosomes of the *G. hirsutum* genome. *In silico* analysis of the PCR results revealed that for 11 of the 12 microsatellite DNA markers used, a single virtual amplicon was synthesized, while for the BNL1694 DNA marker, which is genetically linked to the salt stress resistance trait, two virtual amplicons were synthesized.

Table 1. Chromosomal location of some microsatellite DNA markers

No.	DNA marker	Primer Forward/Reverse	Chromosome	Chromosome position	Amplicon size base pair (bp)
1	BNL3436	AACATAGCCTACCATTGCCG TTGTTTGCCAAATTTGAAGC	D06	2032359	198
2	BNL1167	ACGGCGTTACAAGGCATTAC ACTTAGTTTCTCAAAAAAAAAAATGC	A04	23927090	116
3	BNL3601	TTCCGTTGATGGAAATTGAA ACAAGAATGCGTGTGTCTGC	D04	11572754	175
4	BNL2650	ACCATCCCTCAAAACTCCCT CTGGGTTCTTTTTGGCATGT	A11	124109465	213
5	BNL2569	CAGAGAGCCATTGTGAACGA ATAATGCTAGGGCATGTGGC	A06	2776974	167
6	CGR5732	GCCAAGGTTCATTCCTGAAA TATAGGGCTCATCAGGGTGG	A05	15810127	164
7	CGR5866	GCTTAGGTTGTCATCCTTATTCG CCCTTTGTTCAATTTCTCGTG	D11	63931958	146
8	CGR5883	CGGCGCTACAGGACATAAGA GAGAGCAGGCAGTTGGAGAA	D06	51275683	147
9	DPL0322	AAACCTCGTAGTCATAGGCTCAAA AACTATGCACACAGATTTGGTACG	D01	51621594	198
10	BNL1414	AAAAACCCCTTTCCATCCAT GGGTGTCCTTCCCAAAAATT	A09	77243264	139
11	BNL1694	CGTTTGTTTTCGTGTAACAGG TGGTGGATTCACATCCAAAG	A07/ D07	40367154 / 27936319	240 / 224
12	CIR0246	TTAGGGTTTAGTTGAATGG ATGAACACACGCACG	D02	1054121	167

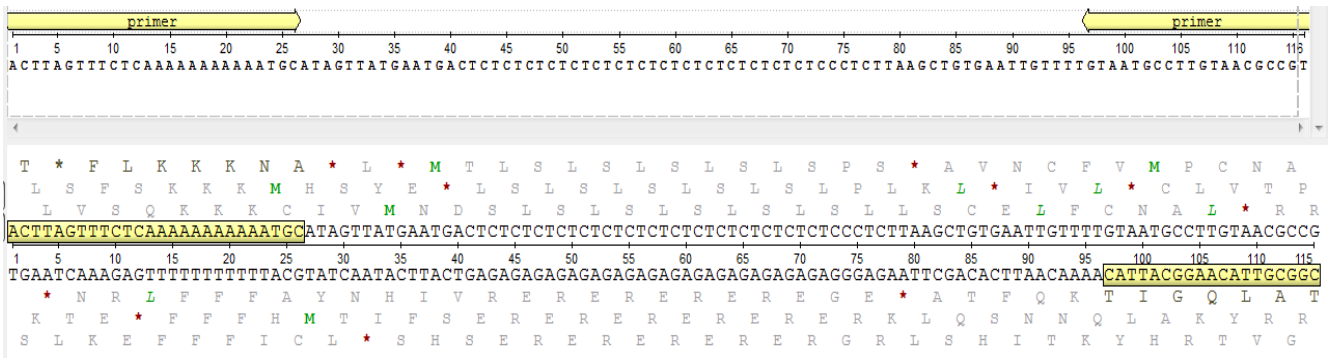


Figure 1. Hybridization of the BNL3436 DNA marker to the *G. hirsutum* genome

Table 2. Upstream and downstream genomic regions of DNA markers

Sr. No.	Marker name	Amplicon size, base pair (bp)	50,000 bp Upstream	50,000 bp Downstream	Associated trait
1	BNL3436	198	1982359	2082556	Number of bolls, fiber length and strength
2	BNL1167	116	23877090	23977205	Fiber quality
3	BNL3601	175	11522754	11622928	Ripening and cell wall thickness
4	BNL2650	213	124059465	124159677	Fusarium wilt resistance
5	BNL2569	167	2726974	2827140	Fusarium wilt resistance
6	CGR5883	147	51225683	51325829	Fiber quality
7	CGR5732	164	15760127	15860290	Fiber strength
8	CGR5866	146	63881958	63982103	Fiber strength
9	DPL0322	198	51571594	51671791	Fiber length
10	BNL1414	139	77193264	77293402	Salt stress tolerance
11	BNL1694	240 / 224	40317154 /27886319	40417393 / 27986542	Salt stress tolerance
12	CIR0246	167	1004121	1104287	Salt stress tolerance

From the perspective of identifying candidate genes and proteins for markers associated with cotton's fiber quality, *Fusarium* wilt resistance, and other valuable agronomic traits, the locus where SSR markers are located in the cotton genome - that is, the region where virtual PCR products of the genome were amplified - was analyzed on both sides of the DNA strand, upstream in the 5' direction

Amplicons were copied, including 50,000 nucleotide pairs upstream and 50,000 nucleotide pairs downstream, for a total of 100,000 nucleotides. Using the nucleotide sequences from this designated genomic region, the AUGUSTUS web application was employed to identify probable genes, their transcript variants, and proteins within the genomic region based on the Arabidopsis thaliana model plant (Table 2).

Subsequently, important genes and proteins specific to the cotton genome were identified by searching for candidate genes using the BLAST search program for putative genes and proteins. During the bioinformatic analysis, 6 candidate genes/proteins were identified from the genomic region of the BNL3436 DNA marker, 5 from the DPL0322 marker region, 4 from the CGR5866 marker region, 2 each from the genomic regions of the BNL2569, CGR5883, CGR5732, BNL1414, BNL1694, and CIR0246 markers, and 1 candidate gene/protein from the BNL3601 marker region. It was found that no candidate genes or transcripts were present in the genomic regions of the BNL1167 and BNL2650 DNA markers (Table 3).

The role of these candidate genes and proteins in various physiological and biochemical processes related to the growth and development of plants, including cotton, was analyzed through a literature review (Supplementary Table 1). In our subsequent analyses, gene primers were designed using bioinformatics programs based on the potential candidate genes found in the genomic region (https://www.ncbi.nlm.nih.gov) (Figures 2, Table 4). Subsequently, conducting molecular studies based on these gene-specific markers, sequencing the genes, and further in-depth molecular investigation of this trait in research samples will create broad opportunities. This will enable the development of new cotton varieties with further improved fiber quality traits that are also resistant to various stress factors and diseases.

Discussion

Several candidate genes involved in various developmental and physiological processes of cotton were identified from the BNL3436 DNA marker region, which is genetically linked to the traits of boll number, fiber length, and fiber strength. In particular, the "Acyl-CoA-binding domain-containing protein 1," identified as a candidate, is an important protein with a size of 10 kDa. Proteins of this ACBP family are present in all eukaryotes and some prokaryotes, and their function and structure have been studied in detail in mammals (humans and mice), plants

(*Oryza sativa*, *Arabidopsis thaliana*, and *Hevea brasiliensis*), and fungi.

These proteins serve to bind medium and long-chain acyl-CoA esters. They promote tolerance to Pb2+ lead ions and their intracellular transport from the root to the shoot; they repress sterol synthesis during embryogenesis and gametophyte development. Most importantly, ACBPs are proteins responsible for lipid transport and play a crucial role in intracellular lipid synthesis, lipid metabolism and remodeling, and gene regulation (8, 9). "Integrin-linked protein kinase 1" (ILK) were discovered in 1996 by Hannigan and his colleagues. ILKs were initially described as serine/threonine-protein kinases but were later proven to be pseudokinases. This enzyme is crucial for cellular proliferation, differentiation, migration, and other processes; it mediates the regulation of cell shape, migration, survival, differentiation, and gene expression (10). Identified through BLAST analysis, the proteins encoding "Zinc finger protein CONSTANS-LIKE 14" play an important role in regulating plant flowering time and, additionally, in providing resistance to drought and salt stress (11). "Somatic embryogenesis receptor-like kinase 1" proteins are present in all organs of the cotton plant and at various developmental stages, with their transcripts being most abundant, particularly in the reproductive organs. Research has shown that these proteins are essential for stamen and pollen cell development (12).

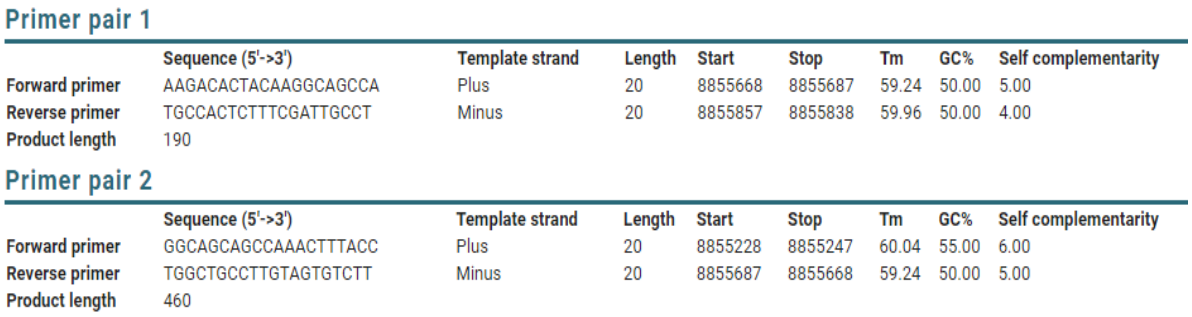
Transcripts of the gene encoding "Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM3" have been found to be essential for male gametophyte development, egg cell specification and function, and are also required for the development of higher-order vascular strands within the leaf, and the interconnected control of leaf shape, size, and symmetry. Furthermore, these proteins regulate the transition of protophloem cells from proliferation to differentiation and participate in the postembryonic growth of the root meristem (13).

Additionally, the "Subtilisin-like protease SBT6.1" proteins identified during the analysis belong to a large family of plant-specific serine peptidases. Research findings indicate that SBT activity is associated with developmental processes and environmental responses. Specifically, 120 and 112 SBTs were identified in the tetraploid species *G. hirsutum* and G. barbadense, while 67 and 69 SBTs were found in the diploid species *G. arboreum* and *G. raimondii*, respectively. These proteins are important in the response to drought stress, and an in-depth study of these genes provides a strong foundation for investigating the drought stress resistance mechanisms of cotton (14). The genes encoding "Senescence-specific cysteine protease SAG39" are located in the genomic region of the DNA marker BNL3601, which is genetically associated with the traits of cotton maturation and cell wall thickness. These genes enable the detoxification of heavy metals, increase yield, ensure the early maturation of cotton, and accelerate the completion of the growing season (15).

Based on a BLAST analysis of the BNL2569 microsatellite marker, which is genetically linked to the Fusarium wilt resistance trait, the probable proteins "MDIS1-interacting receptor like kinase 2" were identified. These proteins regulate plant growth, development, and stress responses. They also participate in various immune reactions, the influx of calcium (Ca2+), providing resistance to several root pathogens, and in processes related to the reception of pollen cells by the pistil (16). Additionally, "Receptor-like protein 9DC3" proteins, identified with 65 in the *G. hirsutum* genome, 23 in *G. arboreum*, and 21 in the *G. raimondii* diploid species (https://www.ncbi.nlm.nih.gov), provide resistance against microbial and fungal pathogens (17). Within the CGR5883 marker region associated with the fiber quality trait, "Pectate lyase" (PEL) proteins were found. Research has identified 53, 42, and 83 putative PEL genes in *G. raimondii* (D5), *G. arboreum* (A2), and *G. hirsutum* (AD1), respectively. These proteins facilitate cell wall softening and remodeling, ensure fiber elongation, contribute to pollen cell development, and allow for the identification of minor differences between long- and short-fiber varieties (18).

Table 3 Genes and transcripts identified in the upstream and downstream genomic regions of DNA markers

Sr. No.	Marker	No. of genes	No. of transcripts	Protein ID
				XP_016717349.2
				XP_016753912.1
				XP_016727140.2
				ADR00582.1
				XP_040964577.1
1	BNL3436	12	13	XP_040947355.1
2	BNL1167	8	8	-
3	BNL3601	5	5	XP_040967102.1
4	BNL2650	10	11	-
				XP_016741832.1
5	CGR5883	11	11	XP_016679142.1
				XP_016672979.2
6	BNL2569	9	9	XP_040955149.1
				XP_016751447.2
				XP_040952778.1
				XP_016667428.1
7	CGR5866	9	10	XP_040957963.1
				XP_016721522.1
				XP_016694420.2
				XP_040967905.1
				XP_040971078.1
8	DPL0322	7	7	XP_040971758.1
				XP_016673260.2
9	CGR5732	8	8	XP_016728883.1
				XP_016687875.2
10	BNL1414	7	7	NP_001314314.1
				XP_016755183.1
11	BNL1694	12	14	XP_016683962.2
				KHG28634.1
12	CIR0246	12	14	GMJ09922.1



Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity
Forward primer	AAGACACTACAAGGCAGCCA	Plus	20	8855668	8855687	59.24	50.00	5.00
Reverse primer	TGCCACTCTTTCGATTGCCT	Minus	20	8855857	8855838	59.96	50.00	4.00
Product length	190							

Primer pair 2

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity
Forward primer	GGCAGCAGCCAAACTTTACC	Plus	20	8855228	8855247	60.04	55.00	6.00
Reverse primer	TGGCTGCCTTGTAGTGTCTT	Minus	20	8855687	8855668	59.24	50.00	5.00
Product length	460							

Figure 2. Selected F/R primer pairs

Table 5. Primer pairs designed based on candidate genes

Sr. No.	Nucleotide sequences (5'->3')	Length	Annealing temperature, °C	GC content (%)	Amplicon size
Subtilisin-like protease SBT6.1					
	F: AGACATGTCGAAAGGCCCAG	20	60.04	55.00	
1	**R:** AGGGATTGCCATTCCTGGTG	20	60.03	55.00	370
	F: GATCAACTCTGCCTGCACCT	20	60.04	55.00	
2	**R:** GATCAACTCTGCCTGCACCT	20	59.97	55.00	734
Receptor-like protein 9DC3					
	F: ACAGTGGTGTCATCTCGCAG	20	60.04	55.00	
1	**R:** AAGATGTGTGGTTCACTAGGAG	22	57.46	45.45	250
	F: TCTCCTAGTGAACCACACATCT	22	58.22	45.45	
2	**R:** CAGTTCCTTGAGAGCGAGTGT	21	60.00	52.38	593
CASP-like protein 4D1					
	F: GTGGATGTTCAGTTCGCACG	20	59.84	55.00	
1	**R:** CAATCCTTAAGGGCGTTCTAGGT	23	60.12	47.83	734
	F: ACCTAGAACGCCCTTAAGGATT	22	58.89	45.45	
2	**R:** ATGGCACCTCCACCAACTTC	20	60.25	55.00	791
Monogalactosyldiacylglycerol synthase, chloroplastic isoform X2					
	F: GTTATCTTTTCGGAGTGGGGA	21	57.38	47.62	
1	**R**: GCAAGGCATGGCATGAAAGT	20	59.75	50.00	179
	F: GATGAGAACCTTGGGGAGCC	20	60.11	60.00	
2	**R:** ATCCCCACTCCGAAAAGATAAC	22	57.85	45.45	612
Putative methyltransferase C9orf114					
	F: GGCAGCAGCCAAACTTTACC	20	60.04	55.00	
1	**R:** TGGCTGCCTTGTAGTGTCTT	20	59.24	50.00	460
	F: AAGACACTACAAGGCAGCCA	20	59.24	50.00	
2	**R:** TGCCACTCTTTCGATTGCCT	20	59.96	50.00	190

Furthermore, during the prediction of candidate genes from this genomic region, a gene named "Ethylene-responsive transcription factor ERF113" was identified, with 4 candidate genes found in G. hirsutum, 1 in *G. arboreum*, and 2 in the diploid species *G. raimondii* (https://www.ncbi.nlm.nih.gov/ gene). The transcription factors of this gene are involved in functions related to stress responses. This gene (cDNA), designated GhERF4, encodes a protein in *G. hirsutum* with a total length of 1061 bp, a molecular weight of 23.5 kDa, and composed of 222 amino acids. GhERF4 transcripts play a significant role in plant development and in responses to abiotic stresses - particularly salinity, cold, and drought - and other environmental factors (19). They are also crucial in transmitting stress tolerance signals specifically related to the regulation of stomatal closure (in response to waterlogging) (20).

Transcription factors known as "Protein FAR-RELATED

SEQUENCE 5" have been identified in the CGR5732 microsatellite marker region, which is genetically linked to the fiber strength trait. These factors play a crucial role in transmitting light signals and regulate plant growth, development, immunity, and defense mechanisms (21). Additionally, a total of 114 orthologs of "Myosin-M heavy chain-like," which were evaluated as candidate genes, have been noted in G. hirsutum, G. barbadense, *G. raimondii*, and *G. arboreum*. As molecular motors, these proteins bind to actin and mediate various physiological processes, such as cell division, movement, migration, and morphology (22).

The presence of genes encoding "ABC transporter B family member 15" was identified for fiber strength trait. These genes and their transcripts belong to a superfamily class of transmembrane proteins that are naturally widespread in organisms. The ABCC (ATP-binding C subfamily) protein is part of a subfamily of the ABC protein family and is a carrier associated with antibiotic resistance. It is localized in the tonoplast and plays a crucial role in pathogenic microbial responses, heavy metal regulation, secondary metabolite transport, and plant growth (23). Furthermore, "Protein MAINTENANCE OF MERISTEMS-like" proteins have been found to play a significant role in promoting flowering and reproductive growth (24), while the "CASP-like protein 4D1" gene, which encodes a membrane domain, plays an important role in cold tolerance. These stress signals are received by receptors, then transmitted and propagated by downstream effectors, ultimately altering the expression of various genes that determine growth, tolerance, and/or survival in response to environmental conditions (25). Transmembrane (TM) proteins located in the plasma membrane are known to have various physiological functions, such as signal reception and recognition through ion and metabolite exchange. It is estimated that Arabidopsis contains approximately 6,500 TM proteins (26).

Additionally, bioinformatic analysis revealed the presence of "Pentatricopeptide repeat-containing protein At1g62930, chloroplastic-like" transcript factors within the designated genome region of the microsatellite marker. Chloroplasts cannot develop normally without the coordinated action of various proteins and signaling connections between the nucleus and the chloroplast genome. It is known that chloroplasts are well-studied as the most important organelles in higher plants. In addition to photosynthesis, chloroplasts are the primary source for the biosynthesis of fatty acids, hormones, amino acids, and other metabolites, as well as for nitrate assimilation (27,28).

Chloroplasts are semi-autonomous organelles whose functions are regulated by proteins encoded by the nuclear genome. The PPR protein family is one of the largest protein families in terrestrial plants, with more than 400 members identified in most plant species (450 in Arabidopsis and 600 in maize) (29,30).

PPR proteins play a crucial role in plant development, particularly in the regulation of RNA within chloroplasts and mitochondria. Each year, more evidence is gathered on the diverse roles of PPR proteins in plants by studying processes such as RNA editing, stabilization, or intron splicing, as well as the transcription and translation of organellar genes (31). Defective or damaged PPR proteins typically lead to negative phenotypes such as delayed embryonic development, pigmentation defects, and abnormal chloroplast biogenesis (pale yellow grana, altered thylakoid formation, and defects in chlorophyll synthesis) (32).

Within the isolated genomic region of the DPL0322 marker, which is associated with the fiber length trait, several candidate genes were identified. These include: "Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 6" (33), which stimulates petal formation and growth; "GDP-mannose 4,6-dehydratase 1" (34), which participates in ensuring the strength of root tissues; "Monogalactosyldiacylglycerol synthase, chloroplastic isoform X2" (35), which plays a crucial role in thylakoid membrane formation, chloroplast biogenesis, and the synthesis of root cell membrane lipids; and "1-phosphatidylinositol-3-phosphate 5-kinase FAB1A isoform X1" (36), which mediates endocytosis, vacuole formation, endomembrane homeostasis, and pollen viability. Additionally, multifunctional "Laccase-12" candidate genes were also found; these genes play an important role in cell elongation, lignification, and pigmentation, and cotton fiber quality is directly dependent on their activity. Genomic analysis of cultivated allotetraploid (G. hirsutum) and its diploid ancestors (*G. arboreum* and *G. raimondii*) confirmed the presence of 84, 44, and 46 laccase genes, respectively. When the expression, enzymatic activity, and biochemical properties of genes associated with fiber development in *G. arboreum* were analyzed, it was found that 40 out of the 44 total laccase genes are involved in various stages of fiber development (37).

In the genomic region where the BNL1414 SSR marker, which is genetically linked to the salt stress resistance trait, is located, the following candidate genes were identified: "L10-interacting MYB domain-containing protein-like" (38), which is involved in pollen development and acts as antiviral immunity, and "MADS-box protein CMB1-like," which controls various developmental processes from root formation to fruit ripening. As a result of studying the genomes of diploid (*G. arboreum* and *G. raimondii*) and tetraploid (G. hirsutum) cotton, 147, 133, and 207 MADS-box genes were described, respectively.

In the search for candidate genes and proteins based on BLAST analysis, transcript variants of the "Kinesin-like protein KIN-7M, chloroplastic" gene were identified in the BNL1694 marker genomic region. These are important for facilitating chloroplast movement and regulating the activity of blue light receptor phototropins (39). Additionally, "Disease resistance protein At4g27190-like" proteins provide resistance to Verticillium wilt, a disease caused by the fungus Verticillium dahliae Kleb. This disease leads to yellowing, wilting, defoliation, and ultimately the death of the cotton plant (40).

Additionally, during the analysis of potential genes and proteins, the presence of genes encoding "Pyridoxal-phosphate-dependent serine hydroxymethyltransferase" was noted in the designated genomic region of the CIR0246 DNA marker. These genes are significant in providing salt tolerance, as well as in leaf growth and the prevention of

chlorosis (41). Furthermore, the "Putative methyltransferase C9orf114" candidate genes serve as an important factor in promoting the development of generative organs and flowering (42).

Conclusion

Thus, based on a virtual analysis of microsatellite markers that exhibited high polymorphism during the study's molecular analysis, 10 SSR DNA markers genetically linked to certain economically valuable traits were identified. During the BLAST analysis of these markers, 29 valuable candidate genes and proteins were identified. In our subsequent analyses, gene primers were designed using bioinformatics programs based on the potential candidate genes found in the genomic region. Subsequently, conducting molecular studies based on these gene-specific markers, sequencing the genes, and further in-depth molecular investigation of this trait in research samples will create broad opportunities. This will enable the development of new cotton varieties with further improved fiber quality traits that are also resistant to various stress factors and diseases.

Supplementary Data:

Supplementary Table 1

Acknowledgement: None
Disclaimer: No
Conflicts of Interest: The authors declare no conflict of interest.
Data availability: Available on reasonable request.
Source of Funding: This research was carried out within the framework of the state budget program for 2025-2030 on the topic “Studying the Abiotic Stress Tolerance of Certain Agricultural Crops Using Modern Biotechnological Approaches” of the Institute of Genetics and Experimental Plant Biology of the Academy of Sciences of the Republic of Uzbekistan.
Ethics approval statement: Not applicable
Authors’ Contribution Statements: F.U.R., B.B.O. conceived and designed the research, M.D.K. conducted analyses using bioinformatics programs. A.K.T., S.K.A., analyzed the data and drafted the manuscript. D.K.E. and F.N.K. managed the project. All authors contributed to manuscript revisions and approved the submitted version.

References

1. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (*Gossypium* hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–30.
2. Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, et al. *Gossypium* barbadense genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci Rep. 2015;5:14139.
3. Meng Q, Gu J, Xu Z, Zhang J, Tang J, Wang A, et al. Comparative analysis of genome sequences of the two cultivated tetraploid cottons, *Gossypium* hirsutum (L.) and G. barbadense (L.). Ind Crops Prod. 2023;196:116471.
4. Khan Z, Khan SH, Ahmed A, Iqbal MU, Mubarik MS, Ghouri MZ, et al. Genome editing in cotton: challenges and opportunities. J Cotton Res. 2023;6(1):3.
5. Okonechnikov K, Golosova O, Fursov M, UGENE Team. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–7.
6. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7:62.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
8. Gallego SF, Sprenger RR, Neess D, Pauling JK, Ejsing CS. Quantitative lipidomics reveals age-dependent perturbations of whole-body lipid metabolism in ACBP-deficient mice. Biochim Biophys Acta Mol Cell Biol Lipids. 2016;1862(2):145–55.
9. Knudsen J, Jensen MV, Hansen JK, Faergeman NJ, Neergaard TB, Gaigg B. Role of acyl-CoA binding protein in acyl-CoA transport, metabolism and cell signaling. Mol Cell Biochem. 1999;192(1-2):95–103.
10. Gorska A, Mazur AJ. Integrin-linked kinase (ILK): The known versus the unknown and perspectives. Cell Mol Life Sci. 2022;79(2):100.
11. Wenqiang Q, Ya Y, Yuying J, et al. Genome-wide analysis of CONSTANS-like genes in cotton. Int J Mol Sci. 2018;19(9):2658.
12. Sha YL, Zhang R, Wang XP, et al. Cloning and characterization of a somatic embryogenesis receptor-like kinase gene in cotton. J Integr Agric. 2012;11(6):898–909.
13. UniProt Consortium. UniProt Entry O65440: LRR receptor-like serine/threonine-protein kinase BAM3 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/O65440/entry
14. Maohua D, Na Z, Yue Z, et al. Genome-wide analysis of the SBT gene family involved in drought tolerance in cotton. Front Plant Sci. 2023;13:1097732.
15. UniProt Consortium. UniProt Entry Q7XWK5: Senescence-specific cysteine protease SAG39 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/Q7XWK5/entry
16. UniProt Consortium. UniProt Entry Q8VZG8: MDIS1-interacting receptor like kinase 2 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/Q8VZG8/entry
17. De Kock MJ, De Wit PJGM. Receptor-like proteins involved in plant disease resistance. Mol Plant Pathol. 2005;6(1):85–97.
18. Huiru S, Pengbo H, Qiang M, Meng Z, Yuan Q, Hengling W, et al. Genome-wide identification and expression analyses of the pectate lyase (PEL) gene family in cotton (*Gossypium* hirsutum L.). BMC Genomics. 2018;19(1):661.
19. Long GJ, Jin YL. Molecular cloning, expression profile and promoter analysis of a novel ethylene-responsive transcription factor gene GhERF4 from cotton

(*Gossypium* hirsutum). Plant Physiol Biochem. 2008;46(1):46–53.
20. UniProt Consortium. UniProt Entry Q9LYU3: Ethylene-responsive transcription factor ERF113 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/Q9LYU3/entry
21. Wang W, Tang W, Ma T, et al. A pair of light signaling factors regulates plant immunity by modulating chlorophyll biosynthesis. J Integr Plant Biol. 2016;58(1):91–103.
22. Chenhui M, Zibo Z, Na W, et al. Genome-wide identification and comparative analysis of myosin gene family in cotton species. Genes (Basel). 2020;11(7):731.
23. Sun N, Xie YF, Wu Y, et al. Genome-wide identification of ABCC gene family and their expression in pigment deposition of cotton fiber. PLoS One. 2021;16(5):e0246649.
24. Chen L, Yan Y, Ke H, et al. SEP-like genes of *Gossypium* hirsutum promote flowering in a concentration-dependent manner. Front Plant Sci. 2022;13:990221.
25. Xiong LM, Schumaker KS, Zhu JK. Cell signaling during cold, drought, and salt stress. Plant Cell. 2002;14(Suppl 1):S165–83.
26. Zhu JH, Dong CH, Zhu JK. Interplay between cold-responsive gene regulation and metabolism. Curr Opin Plant Biol. 2007;10(3):290–5.
27. Sakamoto W, Miyagishima SY, Jarvis P. Chloroplast biogenesis: Control of plastid development, protein import, division and inheritance. Arabidopsis Book. 2008;6:e0110.
28. Waters MT, Langdale JA. The making of a chloroplast. EMBO J. 2009;28(19):2861–73.
29. Lurin C, Andrés C, Aubourg S, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell. 2004;16(8):2089–103.
30. Barkan A, Small I. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol. 2014;65:415–42.
31. Rovira AG, Smith AG. PPR proteins: Orchestrators of organelle RNA metabolism. Physiol Plant. 2019;166(2):451–59.
32. Williams-Carrier R, Brewster C, Belcher SE, et al. The Arabidopsis pentatricopeptide repeat protein LPE1 is required for chloroplast RNA translation. Plant J. 2019;99(1):56–66.
33. UniProt Consortium. UniProt Entry O82268: Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 6 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/O82268/entry
34. Mulichak AM, et al. Structure of the MUR1 GDP-mannose 4,6-dehydratase from Arabidopsis thaliana: Implications for ligand binding and specificity. Biochemistry. 2002;41(52):15578–89.
35. UniProt Consortium. UniProt Entry O82730: Monogalactosyldiacylglycerol synthase, chloroplastic isoform X2 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/O82730/entry
36. UniProt Consortium. UniProt Entry Q9LUM0: 1-phosphatidylinositol-3-phosphate 5-kinase FAB1A isoform X1 [Internet]. UniProt KB; [cited 2026 Aug 28]. Available from: https://www.uniprot.org/uniprotkb/Q9LUM0/entry
37. Balasubramanian VK, Rai KM, Thu SW, Hii MM, Mendu V. Genome-wide identification of multifunctional laccase gene family in cotton (*Gossypium* spp.) and their expression during fiber development. Sci Rep. 2016;6:34309.
38. Yang L, Jiang J, Du ML, et al. A cotton MYB-like transcription factor is involved in pollen development. Plant Cell Physiol. 2013;54(6):893–906.
39. Suetsugu N, Yamada N, Kagawa T, et al. Two kinesin-like proteins mediate actin-based chloroplast movement in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 2010;107(19):8860–5.
40. Sun Q, Jiang HZ, Zhu XY, et al. RNA sequencing analysis of cotton response to Verticillium dahliae infection. BMC Genomics. 2013;14:852.
41. Pragya MP, Jain AJ, Takabe T, et al. Heterologous expression of serine hydroxymethyltransferase-3 from rice confers tolerance to salinity stress in E. coli and Arabidopsis. Front Plant Sci. 2019;10:217.
42. PubMed Central. PMC5381346: Identification of putative methyltransferase C9orf114 function [Internet]. NCBI PMC; [cited 2026 Aug 28]. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5381346

Supplementary Table 1. Putative genes and proteins identified using the NCBI BLAST database

Sr. No	Marker Name	Predicted Gene or Protein	Function
1	BNL3436	*Acyl-CoA-binding domain-containing protein 1*	Lipid transport and metabolism
		Integrin-linked protein kinase 1	Cell proliferation, differentiation, and migration
		Zinc finger protein CONSTANS-LIKE 14	Flowering, drought and salt stress tolerance
		Somatic embryogenesis receptor-like kinase 1 protein	Stamen and pollen cell development
		Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM3	Male gametophyte development, egg cell specification, leaf venation, leaf shape, and size
		Subtilisin-like protease SBT6.1	Drought stress tolerance
2	BNL3601	*Senescence-specific cysteine protease SAG39*	Heavy metal detoxification, accelerated vegetation
3	BNL2569	*MDIS1-interacting receptor-like kinase 2*	Regulation of growth and development, resistance to root pathogens
		Receptor-like protein 9DC3	Resistance to bacterial and fungal pathogens
4	CGR5883	*Pectate lyase*	Cell wall formation, ensuring fiber elongation, pollen cell development
		Ethylene-responsive transcription factor ERF113	Tolerance to salinity, cold, drought, and waterlogging stress
5	CGR5732	*Protein FAR-RELATED SEQUENCE 5*	Light signal transduction, growth, development, immunity, and defense
		Myosin-M heavy chain-like	Cell division, movement, migration, and morphology
6	CGR5866	*ABC transporter B family member 15*	Transport of heavy metals and secondary metabolites
		Protein MAINTENANCE OF MERISTEMS-like	Promotion of generative organ development and flowering
		Pentatricopeptide repeat-containing protein At1g62930, chloroplastic-like	Plant development, RNA stabilization in chloroplasts and mitochondria
		CASP-like protein 4D1	Cold tolerance

7	DPL0322	*Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 6*	Promotion of petal formation and growth
		GDP-mannose 4,6-dehydratase 1	Root tissue strength
		Laccase-12	Cell elongation, lignification, and pigmentation
		Monogalactosyldiacylglycerol synthase, chloroplastic isoform X2	Chloroplast biogenesis, synthesis of root cell membrane lipids
		1-phosphatidylinositol-3-phosphate 5-kinase FAB1A isoform X1	Endocytosis, vacuole formation, maintenance of membrane homeostasis
8	BNL1414	*L10-interacting MYB domain-containing protein-like*	Pollen development, antiviral immunity
		MADS-box protein CMB1-like	Control of various developmental processes from root formation to fruit ripening
9	BNL1694	*Kinesin-like protein KIN-7M, chloroplastic*	Ensuring chloroplast movement
		Disease resistance protein At4g27190-like	Verticillium wilt resistance
10	CIR0246	*Pyridoxal-phosphate-dependent serine hydroxymethyltransferase*	Salinity tolerance
		Putative methyltransferase C9orf114	Stimulation of generative organ development and flowering