IRSTI 34.31.19 https://doi.org/10.52081/bkaku.2025.v75.i4.281

GENOME SEQUENCE RESOURCES OF UPLAND COTTON (*GOSSYPIUM HIRSUTUM*) PROVIDES MOLECULAR STRUCTURE FOR ADVANCED BREEDING EFFORTS

Amangeldyyeva N.K.[3], Master's student 2 year 7M05107 Biology
aimeshka_1609@inbox.ru, https://orcid.org/0009-0000-5101-1250
Orken A.[1,3], PhD student 1 year 8D05107 Biology
orkena23@gmail.com, https://orcid.org/0009-0006-8158-9305
Zhumabay N.B.[1], PhD student 1 year 8D08100 Agronomy
nurbek.zhumabay10@gmail.com, https://orcid.org/0009-0001-0033-5834
Manabayeva Sh.A.[1,3], Professor, Candidate of Biological sciences
manabayeva@biocenter.kz, https://orcid.org/0000-0001-7884-1713
Makhmadzhanov S.P.[2], Associate professor, candidate of Agricultural Sciences
max_s1969@mail.ru, https://orcid.org/0000-0001-5623-0591
Tussipkan D.[1*], Associate professor, PhD
tdilnur@mail.ru, https://orcid.org/0000-0003-1337-2834

[1]*National center for biotechnology, Korgalzhyn hwy. 13/5, Astana, Kazakhstan*
[2]*Agricultural Experimental Station of Cotton and Melon Growing, Atakent, Kazakhstan*
[3]*L.N. Gumilyov Eurasian National University, Satpayev st. 2, Astana, Kazakhstan*

Annotation. Advances in genome sequencing over the past decade have revolutionized our understanding of *Gossypium hirsutum* by transforming fragmented short-read assemblies into near-complete, chromosome-scale genomic maps. In this review, genome assemblies of key *G. hirsutum* cultivars, beginning with the reference line TM-1, which provided the foundation for the earliest high-quality genome assembly, were examined. Then, additional cultivars, such as CRI-12 and ZM24, and several elite breeding lines, whose assemblies have refined knowledge of genome structure, gene content, and structural variation, were considered. These genomic resources offer critical insights into the genetic basis of traits such as fiber quality, yield, and stress tolerance. Next, recent advances in genomic tools and breeding strategies were considered. Particular attention was given to genome-wide association studies (GWAS), high-density SNP arrays, and transcriptomic profiling, which accelerate trait mapping and candidate gene identification. By synthesizing progress in genome assemblies and molecular breeding, this review underscores the central role of genomic resources in developing resilient, high-yielding cultivars to meet global agricultural and industrial challenges.

Keywords: Upland cotton *(Gossypium hirsutum)*, whole genome sequencing, genome assembly, pangenome, GWAS, molecular breeding, structural variation

Introduction. *Gossypium hirsutum* L., commonly known as Upland cotton, is the most widely cultivated cotton species worldwide, accounting for approximately 95% of global cotton fiber production [1]. Due to its broad ecological adaptability, high yield potential, and economic importance, *G. hirsutum* has become a central focus of modern breeding programs aimed at improving fiber quality, productivity, and tolerance to biotic and abiotic stresses [2]. However, traditional breeding approaches, which rely primarily on phenotypic selection and long-term field evaluation, are constrained by the complexity of its allotetraploid genome (AADD, 2n = 4x = 52), high heterozygosity, and substantial structural variation [3].

The rapid development of high-throughput sequencing technologies over the last decade has dramatically transformed the landscape of cotton genomics. Early genome sequencing attempts produced fragmented and incomplete assemblies, limiting in-depth molecular analyses. The advent of long-read sequencing platforms such as PacBio and Oxford Nanopore, combined with Hi-C–

based chromosomal scaffolding and hybrid assembly strategies, enabled a transition from draft genomes to highly contiguous, chromosome-level assemblies. The reference genome of the TM-1 line laid the foundation for comprehensive genomic studies, while subsequent sequencing of additional cultivars such as ZM24, CRI-12, ZM113, BARBREN-713, NDM8, and Yuanmian11 greatly expanded our understanding of genomic diversity, structural rearrangements, and functional organization within *G. hirsutum*.

Advancements in genomic research on *G. hirsutum* provide a fundamental framework for next-generation breeding technologies by enabling a deeper understanding of the genetic architecture underlying economically important traits. This review aims to systematize current knowledge on Upland cotton genome sequencing efforts, characterize major cultivars and their genomic features, and evaluate the role of these resources in developing highly productive, stress-resistant, and adaptable cotton varieties.

Materials and methods. This review focuses on *G. hirsutum* as the primary research material. Particular attention was given to *G. hirsutum* cultivars and related genomic resources that contribute to understanding cotton evolution, genome structure, and breeding potential.

The literatures for this review was obtained from several major scientific databases, including the Web of Science Core Collection, PubMed, Google Scholar and the National Center for Biotechnology Information. A total of more than 41 publications were examined, comprising research articles, review papers, books, and book chapters published in leading international and regional journals. The search covered the period from 2012 to 2025 to ensure the inclusion of both foundational and recent advances in cotton genomics.

The search strategy utilized a combination of topical keywords and controlled vocabulary terms, including: *Gossypium* L., cotton genetics and genomics, genome structure and variation, SNP arrays, GWAS, QTL mapping. Boolean operators ("AND", "OR") and filters by publication year, article type were applied to refine search results. Reference lists from key papers were manually screened to identify additional relevant studies. Microsoft Office Excel 2021 was used for quantitative analysis of the literature and made table.

Genome assemblies of *Gossypium hirsutum* cultivars. With the advancement of molecular biology and genomics, since 2015 large-scale studies have been initiated to perform detailed sequencing and annotation of the *G. hirsutum* genome. This stage marked a significant shift from phylogenetic reconstructions to an in-depth molecular analysis of genome structure and function.

The **Texas Marker 1 (TM-1)** line of *G. hirsutum*, developed through over 50 generations of self-pollination, has been established as the **genetic standard** for Upland cotton. Owing to its high homozygosity, TM-1 serves as a key reference in genomic and breeding studies.

The sequencing and assembly of the *G. hirsutum* genome has undergone a remarkable transformation over the past decade, evolving from highly fragmented drafts to a nearly gapless, chromosome-level reference. The first widely cited assembly, presented by Saski et al. (2017) [14] using Illumina MiSeq, captured only 169.3 Mb—several times smaller than the actual genome size of approximately 2.3 Gb. It included just four chromosomes and had a scaffold N50 of 48 Mb, reflecting significant fragmentation. A subsequent effort by Li et al. (2015) [15] employed Illumina HiSeq 2000, expanding the assembly to 2.2 Gb and recovering all 26 chromosomes, but the genome remained heavily fragmented into 9,146 scaffolds, and the longest chromosome reached only about 117 Mb, still below the lengths observed in high-quality assemblies. The transition to long-read sequencing marked a turning point. Yang et al. (2019) [16] used PacBio Sequel to produce the first chromosome-scale assembly, reducing the scaffold count to 599 and achieving an N50 of 96.7 Mb. For the first time, chromosome length extremes were clearly defined, with the longest chromosome (Ah06) measuring approximately 128.2 Mb and the shortest (Dh03) about 54.9 Mb. Chen et al. (2020) [17] advanced these results using PacBio RSII, raising the N50 to 108.1 Mb and producing a highly contiguous genome of 2.3 Gb that included more than 75,000 protein-coding genes. The current gold standard is the assembly by Zhang (2024) [17], which combined Oxford Nanopore PromethION with PacBio Sequel. This effort generated just 26 contigs one per chromosome with an

N50 of 108.2 Mb and an L50 of 10, representing minimal fragmentation and near-complete continuity (Table 1). The longest and shortest chromosomes remained consistent with earlier long-read assemblies, underscoring the stability of these genomic features. This steady improvement from limited Illumina-based drafts to an almost perfect hybrid assembly highlights the rapid progress in sequencing and assembly methods.

While the TM-1 cultivar has served as the primary reference genome for *Gossypium hirsutum*, sequencing additional cultivars is essential to capture the broader genetic diversity of upland cotton. Different cultivars exhibit unique adaptations, fiber qualities, and stress tolerances that are often the result of both natural selection and targeted breeding efforts. By generating and analyzing high-quality genome assemblies for these varieties, researchers can identify structural variations, gene content differences, and novel alleles absent from the TM-1 reference. Such comparative genomic studies provide critical insights into trait evolution, facilitate the discovery of candidate genes for breeding programs, and help refine the pangenome of *G. hirsutum*. This section focuses on recent sequencing efforts involving multiple cultivars, detailing their assembly strategies, genome statistics, and potential implications for cotton improvement.

The *ZM24* cultivar was characterized by Yang *et al.* (2019) at the Chinese Academy of Agricultural Sciences (CAAS). Genomic DNA was extracted from seedlings, with Zhongzhimian No. 2 as the parental line. High-molecular-weight DNA was sequenced on the PacBio Sequel platform, and assembly was performed using Canu v1.5. The assembled genome size was estimated at 2.3 Gb, with an average sequencing depth of 1.0× and a GC content of 34.5%. Among the assembled chromosomes, the largest was A06 (121,900,087 bp) and the smallest was D09 (50,656,052 bp). The Zhongzhimian No. 2 cultivar was characterized by Li, Zhang *et al.* (2023) at the Institute of Plant Protection, CAAS. DNA was extracted from seedling leaves, with Zhongmian 113 (ZM113) as the parental line. Sequencing was conducted using PacBio Sequel and Illumina platforms, and assembly was carried out with NextDenovo v2.3.0. The genome size was 2.3 Gb, with 125.0× coverage and a GC content of 34.5%. The largest chromosome was A06 (127,742,751 bp) and the smallest was D03 (53,820,448 bp).

The Zhongmian 113 (ZM113) cultivar was described by Hu, G. and Wang, Z. (2024) at the Institute of Cotton Research, CAAS. DNA was isolated from seedling leaves, with BARBREN-713 as the parental line. Sequencing employed Oxford Nanopore GridION, PacBio Sequel, Illumina NextSeq, and Oxford Nanopore platforms. Genome assembly utilized Hifiasm v0.19.3-r572, Hi-C-Pro v2.8.1, LACHESIS, NextDenovo v2.3.1, and NextPolish v1.3.0. The final assembly size was 2.3 Gb, with 150.0× coverage and a GC content of 34.5%. The largest chromosome was A06 (128,015,698 bp) and the smallest was D03 (54,059,521 bp).

The BARBREN-713 cultivar was analyzed by Perkin *et al.* (2021) at the USDA. DNA was extracted from entire seedlings, with Barbren-713-32-30 as the parental line. Sequencing was performed on the PacBio Sequel platform, and assembly was done using Hifiasm v0.13-r307. The genome size was 2.4 Gb, with 49.0× coverage and a GC content of 35.5%. The largest chromosome was A06 (127,477,704 bp) and the smallest was D03 (54,675,798 bp).

The Barbren-713-32-30 cultivar was reported by Perkin, Bell *et al.* (2021) at the USDA. DNA was extracted from entire seedlings, with NDM8 as the parental line. Sequencing was conducted using PacBio Sequel, and assembly employed Hifiasm v0.13-r307. The genome size was 2.4 Gb, with 36.0× coverage and a GC content of 35.0%. The largest chromosome was A06 (127,360,497 bp) and the smallest was D03 (54,589,988 bp).

The NDM8 cultivar was described by Ma (2021) at Hebei Agricultural University. Genomic DNA was extracted from leaves of two-week-old seedlings. Sequencing was performed on PacBio RSII and Illumina HiSeq platforms, and genome assembly was conducted using FALCON v1.0. The assembled genome size was 2.3 Gb, with a sequencing depth of 89.6× and a GC content of 34.5%. The largest chromosome was A06 (119,073,090 bp) and the smallest was D09 (54,703,076 bp).

Table 1 – Genomic characteristics of *G. hirsutum* cultivars

Cultivar	ZM24	Zhongzhimian No.2	Zhongmian 113 (ZM113)	BARBREN-713	Barbren-713-32-30	NDM8
1	**2**	**3**	**4**	**5**	**6**	**7**
Development stage	not collected	seedling	seedling	seedling	seedling	two weeks
Tissue	seedling	young leaves	young leaves	entire plant	entire plant	young leaves
Geographic location	China	China: Henan	China	USA: College Station	USA: College Station	China
GenBank	CM017385.1-CM017410.1	CM045129.1-CM045154.1	CM104155.1-CM104180.1	CM038291.1-CM038316.1	CM038317.1-CM038342.1	CM032202.1-CM032227.1
Submitter	Chinese Academy of Agricultural Sciences	Chinese Academy of Agricultural Sciences	Institute of cotton research of CAAS	USDA	USDA	Hebei Agricultural University
Date	Jul 29, 2019	Aug 9, 2022	Jan 30, 2025	Jan 13, 2022	Jan 13, 2022	Jun 24, 2021
Assembly type	haploid	haploid	haploid	haploid	haploid	haploid
Assembly level	Chromosome	Chromosome	Complete Genome	Chromosome	Chromosome	Chromosome
Sequencing technology	PacBio Sequel	PacBio Sequel	PacBio Sequel	PacBio Sequel	PacBio Sequel	PacBio RSII
Assembly method	Canu v. 1.5	NextDenovo v. 2.3.0	LACHESIS v. NextDenovo v. v2.3.1	hifiasm v. 0.13-r307	hifiasm v. 0.13-r307	FALCON v. 1.0
Genome size	2.3 Gb	2.3 Gb	2.3 Gb	2.4 Gb	2.4 Gb	2.3 Gb
Total ungap-ped length	2.3 Gb	2.3 Gb	2.3 Gb	2.4 Gb	2.4 Gb	2.3 Gb
Genome Coverage	1.0x	125.0x	150.0x	49.0x	36.0x	89.6x
GC percent	34.5	34.5	34.5	35.5	35	34.5
chr A01	112 577 161	121043734	120 628 686	119 731 888	119 383 478	119073090
chr A02	104 266 397	109 264 372	108 366 301	107964587	108 284 803	107674 087
chr A03	105 393 268	114 342 330	113 591 098	113 515 339	113723839	113072 090
chrA04	80 868 428	89 470 856	89 294 814	89135339	89 061 088	88948101
chr A05	107 448 258	112 971 952	114 104 290	112749629	113294159	112049794
chr A06	121 900 087	127 742 751	128 015 698	127477704	127 360 497	127097621
chr A07	93 248 268	99 241 639	99 024 949	98 310 344	98365699	98540322
chr A08	121 232 165	127 282 784	126 367 618	126 075 804	125486242	126523771
chr A09	79 884 170	86 366 132	88 768 367	88 669 301	88519516	83270209
chr A10	109 403 145	118 750 516	117 804 371	118 207 260	118071405	118226 759
chr A11	115914562	125 156 326	123843250	124 361 224	124 380 747	123047 965
chr A12	102340594	109 904 319	109 204 060	109 205 764	109263233	109222 716
chr A13	107 345 356	122 444 943	112 013 154	112615166	111982279	111309922
chr D01	61090159	65 136 658	65 763 041	66207413	66293935	66018313
chr D02	68262679	73 471 370	73 180 589	74339275	73 927 771	73 080 204
chr D03	53049694	53 820 448	54059521	54675798	54 589 988	54703076
chr D04	54327443	58 609 752	59 504 096	59 435 069	59 483 441	58732198
chr D05	62317630	71 744 611	67 228 931	67 264 888	67 944 199	67339501
chr D06	63 144 374	68 098 790	66 462 893	66728972	66 724 857	66853151
chr D07	55165807	59 767 106	61 139 987	61 060 688	61154651	61 153 439
chr D08	66058605	70425916	71805435	71947194	71 994 581	69886574
chr D09	50656052	54 158 153	57 310 315	55 172 275	55 950 205	55028699
chr D10	64331360	69 384 279	69 073 422	68 797 300	69 022 009	69313209
chrD11	69233432	75 329 889	74 034 940	74325253	73930370	73297438
chr D12	60589447	62 856 854	63 419 764	63320231	63382697	63248683
chrD13	60 180 343	66 343 539	65 051 914	64760530	64960949	65224646
References	[16]	[18]	[17]	[20]	[20]	[21]

continuation of the table 1

Cultivar	JBM	Yuanmian11	YZ1	CSX8308	PSC355	UA48
8	9	10	11	12	13	14
Development stage	young seedling	seedling	not applicable	leaf area and canopy development	Whole seedling	leaf area and canopy development
Tissue	young leaves	young leaves	young leaves	young leaves	not applicable	young leaves
Geographic location	China: Anyang	China: Xinjiang	China: Hubei	USA: South Carolina	USA: Texas	USA:South Carolina
GenBank	CM045499.1-CM045524.1	CM069105.1-CM069130.1	CM074220.1-CM074245.1	CM076542.1-CM076567.1	CM082153.1CM082178.1	CM082236.1-CM082261.1
Submitter	The institute of Cotton Research of CAASc	Ningxia Academy of Agriculture and Forestry Sciences	Huazhong Agricultural University	HudsonAlpha Institute for Biotechnology	USDA-ARS	HudsonAlpha Institute for Biotechnology
Date	Aug 17, 2022	Jan 12, 2024	Mar 18, 2024	Apr 24, 2024	Jul 30, 2024	Jul 30, 2024
Assembly type	haploid	haploid	haploid	haploid	haploid	haploid
Assembly level	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome	Chromosome
Sequencing technology	PacBio Sequel	PacBio HiFi	PacBio Sequel	PacBio Sequel	PacBio Sequel	PacBio Sequel
Assembly method	FALCON v. 1	hifiasm v. 0.16.1	hifiasm v. 0.16.1	MECAT v. 1.3	hifiasm v. 03/18/2022	MECAT v. 1.4
Genome size	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb
Total ungap-ped length	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb	2.3 Gb
Genome Coverage	128.0x	40.0x	18.0x	179.0x	40.9x	161.0x
GC percent	34.5	34.5	34.5	34.5	34.5	34.5
chr A01	119 487501	119 503 960	120 249496	119 381165	119 474 596	117112808
chr A02	106956 574	108475289	108211375	107188 849	108 240 954	105664978
chr A03	113005158	113 458 299	113500973	112124 782	113 544 907	111010592
chrA04	88 659 587	88 956 553	89 337 491	88712028	90 025 657	88 451 855
chr A05	110920663	112 953 411	112613041	109991 617	112 537 697	109917801
chr A06	129207784	128 147 158	128075698	125428 350	128169327	126621437
chr A07	98851549	98 581 300	98 446 870	98 241 532	98 278 401	97 335 875
chr A08	125754566	126 444 814	126280128	125201 531	126 651 951	125748362
chr A09	84538544	87 981 018	87 408 150	84 008 366	87 012 727	83 019 028
chr A10	117587747	117979994	118248507	117248 284	118 101 055	117182076
chr A11	123337410	124 292 941	124313 730	122992942	124 146 241	122100811
chr A12	108770671	109 357 382	109210752	107737407	109 175 953	108042066
chr A13	111406123	111 981 636	112156509	111082 403	112117856	111143903
chr D01	67 005 298	65 874 432	65713259	66904707	66 201 905	65 558 898
chr D02	73787793	73 423 024	73570973	72562124	73 250 298	71 457 850
chr D03	55080092	54 737 121	54558231	54 602 881	54 720 647	54539853
chr D04	58826415	60 261 276	59381772	58397551	59 665 483	57408129
chr D05	66072703	67 605 726	66 914 693	67136090	65 908 016	65 327 979
chr D06	66 737 491	66 763 680	66 572 821	66 247 743	66 868 618	66329078
chr D07	62 054 521	61 456 119	60 679 514	59 643 172	61 041 968	58286762
chr D08	70 204 606	71 743 061	71821552	69 655 482	71 775 203	68510787
chr D09	54809226	56 121 625	55760029	54 917 242	56087791	53 813 707
chr D10	68 788 203	68 821 615	68879443	67837702	68 597 512	68 178 173
chrD11	73 120 872	73 003 245	73541740	73488099	74015571	72 849 196
chr D12	63032241	63 445 998	63455446	63 245 180	63 565 837	62835596
chrD13	65063187	65063231	65 100 670	65 235 532	65 155 209	64 558 943
References	[22]	[23]	[24]	[25]	[26]	[27]

The JBM cultivar was characterized by Peng *et al.* (2021) at the Institute of Cotton Research, Chinese Academy of Agricultural Sciences. Genomic DNA was extracted from leaf tissue of young seedlings, sequenced using the PacBio Sequel platform, and assembled with FALCON v1.0. The genome size was 2.3 Gb, with 128.0× coverage and a GC content of 34.5%. Chromosome A06 was the largest (119,487,501 bp) and D09 the smallest (55,080,092 bp).

The Yuanmian 11 cultivar was investigated by Wang, Liang *et al.* (2023) at the Ningxia Academy of Agriculture and Forestry Sciences. DNA was extracted from seedling leaf tissue, sequenced using PacBio HiFi, and assembled with hifiasm v0.16.1. The genome size was 2.3 Gb, with 40.0× coverage and a GC content of 34.5%. The largest chromosome was A06 (119,503,960 bp) and the smallest was D09 (54,737,121 bp).

The YZ1 cultivar was studied by Xu (2024) at Huazhong Agricultural University. Genomic DNA was extracted from leaf tissue, with no developmental stage specified. Sequencing was performed on PacBio Sequel, Oxford Nanopore PromethION, and Illumina HiSeq platforms, and the assembly was generated using hifiasm v0.16.1. The final genome size was 2.3 Gb, with 18.0× coverage and a GC content of 34.5%. Chromosome A06 measured 120,249,496 bp, while D09 measured 54,558,231 bp.

The CSX8308 cultivar was characterized by Stiller *et al.* (2024) at the HudsonAlpha Institute for Biotechnology. DNA was extracted from young leaves collected during the leaf area and canopy development stage. Sequencing was conducted on PacBio Sequel II and Illumina NovaSeq platforms, with assembly performed using MECAT v1.3. The genome size was 2.3 Gb, with 179.0× coverage and a GC content of 34.5%. Chromosome A06 was the largest (119,381,165 bp) and D09 the smallest (54,602,881 bp).

The PSC355 cultivar was described by Cohen, Perkin *et al.* (2024) at the USDA-ARS. Genomic DNA source was not specified. Sequencing was performed using the PacBio Sequel platform, and assembly employed hifiasm (version dated 03/18/2022). The genome size was 2.3 Gb, with 40.9× coverage and a GC content of 34.5%. The largest chromosome was A06 (119,474,596 bp) and the smallest was D09 (54,720,647 bp).

The UA48 cultivar was analyzed by Bourland *et al.* (2024) at the HudsonAlpha Institute for Biotechnology. DNA was extracted from young leaves at the leaf area and canopy development stage, sequenced using PacBio Sequel II and Illumina NovaSeq platforms, and assembled with MECAT v1.4. The genome size was 2.3 Gb, with 161.0× coverage and a GC content of 34.5%. Chromosome A06 measured 117,112,808 bp, and D09 measured 54,539,853 bp.

Comparative analysis of these genome assemblies confirms the conserved chromosomal organization of upland cotton while highlighting subtle differences associated with cultivar-specific adaptations. The table 1 summarizes 13 haploid-level genome assemblies of upland cotton cultivars from China and the USA, generated between 2019 and 2025. These assemblies were produced using advanced sequencing platforms, mainly PacBio Sequel and Sequel II, often complemented with Illumina or Oxford Nanopore technologies, and assembled with tools such as Canu, FALCON, MECAT, hifiasm, and NextDenovo. Genome sizes are consistent at 2.3–2.4 Gb with a GC content of ~34.5–35.5%. The samples were collected at different developmental stages and tissues, reflecting diverse experimental designs Chromosome A06 was invariably the largest, while D09 (or D03 in some cases) was the smallest. Variation in sequencing technologies, assembly algorithms, and parental lineages reflects both advances in genomic methodologies and the distinct breeding histories of these cultivars. These high-quality assemblies substantially enrich cotton genomic resources, supporting pangenome development, structural variant detection, and the identification of alleles linked to agronomic performance, stress tolerance, and fiber quality—ultimately informing targeted breeding strategies for *G. hirsutum* improvement (Table 1).

Advances in genomic tools and breeding strategies for *G. hirsutum*. Whole Genome Sequencing (WGS) is a comprehensive genomic analysis technique that determines the complete DNA sequence of an organism's genome in a single process. By providing base-by-base coverage, WGS allows the identification of virtually all types of genetic variation, including single nucleotide

polymorphisms (SNPs), insertions and deletions (indels), structural variants, and copy number changes [29]. For crop species such as *G. hirsutum*, WGS enables the discovery of millions of SNPs across the genome, which can subsequently be filtered and curated to design high-density genotyping arrays [30]. WGS provides flexibility for exploring novel loci, detecting rare variants, and identifying genomic regions under selection. This makes WGS not only a source of high-quality marker datasets but also a strategic platform for developing resources such as the CottonSNP63K and CottonSNP80K arrays.

CottonSNP63K array

Single nucleotide polymorphisms (SNPs) are the most common type of variation in plant genomes. They are evenly distributed across the genome, bi-allelic, and codominant, which makes them a valuable tool for genotyping, quantitative trait locus (QTL) mapping, and marker-assisted selection. A breakthrough came with the development of the CottonSNP63K array, designed using genomic libraries (RAD-seq and RNA-seq) from 11 cotton lines and interspecific sequencing data. From 1.2 million candidate SNPs, 45,000 interspecific and approximately 18,000 intraspecific markers were selected. The panel proved useful for trait mapping and diversity analysis but had limited coverage of intraspecific polymorphisms within *G. hirsutum* [31].

If the CottonSNP63K array became the first universal tool for cotton genotyping, the CottonSNP80K array significantly enhanced intraspecific resolution, making it possible to conduct fine-scale genetic analyses within *G. hirsutum*. Together, these arrays marked the transition from a limited set of markers to high-density platforms, opening new horizons for functional genomics and cotton breeding.

CottonSNP80K array

The next step was the creation of the CottonSNP80K array, specifically optimized for intraspecific diversity. Its development was based on the reference genome sequence of the TM-1 line and resequencing data from 100 *G. hirsutum* cultivars. From 1.37 million candidate SNPs, those passing filters for minor allele frequency (MAF > 0.1), flanking sequence quality, and Illumina design scores were narrowed down to 82,259. The final array, implemented on the Illumina Infinium platform, contained 77,774 markers evenly distributed across the genome, averaging one SNP every ~25 kb.

This technological advancement enabled highly accurate genotyping with reproducibility rates of up to 100% and more than 95% concordance with whole-genome sequencing data, facilitated the mapping of important agronomic traits through GWAS, traced the origin of modern cultivars and the contribution of landraces and introduced forms, accelerated the identification of valuable alleles for breeding programs, and allowed for the verification of cultivar identity and the detection of heterozygosity in F1 hybrids [32-33].

Axiom® Cotton Genotyping Array

There is another array, the Axiom® Cotton Genotyping Array, which is relatively new in this field. It was developed by Affymetrix in collaboration with the National Botanical Research Institute, India, and represents one of the most comprehensive tools for high-density SNP genotyping in cotton. The array includes 35,550 markers identified in *G.hirsutum* and *G. barbadense* (Pima cotton), the two species that account for the majority of global cotton production. Cotton is crop with a narrow germplasm base, making SNP discovery and genotyping particularly challenging due to the low frequency of polymorphisms. Traditional approaches such as genotyping-by-sequencing (GBS) have proven inefficient and costly for this genome type. The Axiom® Cotton Genotyping Array overcomes these obstacles by providing robust, reproducible, and cost-effective genotyping results.

The markers on the array were selected from multiple sources, including gene-enriched genomic sequences of *G. hirsutum*, genome reduction approaches based on restriction site conservation (GR-RSC), and interspecific assemblies of *G. hirsutum* and *G. barbadense*. Importantly, marker discovery was performed using diverse accessions representing both cultivated and wild cotton forms, as well as lines with contrasting fiber qualities, ensuring the array's

relevance for both diversity studies and breeding applications [34]. This platform is suitable for a wide range of applications: genome-wide association studies (GWAS), QTL mapping, marker-trait association, molecular breeding, and SNP discovery. It provides reliable data for analyzing complex traits of economic importance and for accelerating the development of new cotton varieties through marker-assisted selection.

GWAS as a tool for dissecting complex traits in plants

Genome-wide association studies (GWAS) are a powerful approach used to identify genetic loci associated with phenotypic traits by scanning the entire genome for statistically significant marker–trait associations. Unlike traditional linkage mapping, which relies on bi-parental populations, GWAS leverages natural genetic variation in diverse germplasm collections, offering higher resolution for locating causal genes. In plants, GWAS has become an essential tool for dissecting the genetic architecture of complex traits, such as yield, quality, and stress tolerance.

The first genome-wide physical map of *G. hirsutum* was constructed by Zhang et al. (2012) using a BIBAC-based approach. They developed a comprehensive physical map consisting of 3,450 BIBAC contigs with an N₅₀ size of 863 kb, spanning approximately 2,244 Mb — covering nearly the entire genome (≈ 92.6%) and sorting contigs by their A- and D-subgenome origins [40]. This map, annotated with ~10,000 BIBAC-end sequences (BESs) (one per ~250 kb), provided a crucial platform for fine mapping, gene/QTL cloning, and sequencing efforts [35].

Before that, in 2010, a draft physical map of the diploid progenitor species *Gossypium raimondii* (D-genome) was constructed by Lin et al., integrating overgo hybridization probes, agarose fingerprinting, and high-information-content fingerprinting (HICF). This connected 1,585 contigs to a consensus cotton genetic map, anchoring them also to *Arabidopsis* and grapevine genomes, thereby aiding assembly and comparative genomics [36].

More recent advances include the development of subgenome-anchored physical maps for upland cotton. In 2017, Saski et al. constructed BAC libraries and generated a de novo whole-genome physical map, partitioning it into A- and D-subgenomes using BES alignment, FISH validation, and SNP genetic mapping. This approach delivered the first subgenome-anchored physical maps and advanced the road toward obtaining a reference-grade genome assembly.

One of the earliest systematic applications of GWAS for investigating drought tolerance in *G. hirsutum* was carried out by Hou et al. (2018) [37], who used a natural panel of cotton genotypes phenotyped under both water-deficit and well-watered conditions. Their analysis revealed significant SNP associations with traits linked to drought adaptation, including loci near genes involved in abscisic acid (ABA) signaling, stomatal regulation, and root system development, laying the groundwork for molecular marker–assisted selection in breeding for stress tolerance.

Building on this foundation, Guo et al. (2022) [38] applied the high-density CottonSNP80K array to a population of *G. hirsutum* races, identifying novel loci associated with physiological traits such as water-use efficiency and photosynthetic performance under drought conditions. By integrating GWAS findings with transcriptomic data, they pinpointed transcription factors such as DREB, NAC, and bZIP, along with aquaporin genes essential for maintaining cellular turgor.

More recently, Sun et al. (2023) [39] conducted GWAS on 150 *G. hirsutum* genotypes grown under normal irrigation and water-limited environments, discovering SNP markers strongly correlated with yield components and root morphology under drought stress. These markers were recommended for use in marker-assisted and genomic selection to accelerate the introgression of drought-tolerance traits into elite lines.

Collectively, these studies demonstrate that GWAS, especially when paired with high-density SNP arrays, offers a powerful means of uncovering genes and alleles underlying drought response, providing valuable molecular tools for breeding and advancing the understanding of the mechanisms driving cotton's adaptation to water stress, thereby bridging fundamental genomics and applied crop improvement to enable the development of climate-resilient cultivars.

Meanwhile, wild species of *Gossypium* are recognized as invaluable reservoirs of genetic diversity for breeding. They possess traits such as disease resistance and abiotic stress tolerance that

can be introgressed into *G. hirsutum*. Efforts to utilize wild species like *G. mustelinum* have shown improvements in fiber quality through backcrossing generation trials, although reproductive barriers such as hybrid breakdown and daylength sensitivity pose challenges [40]. Reviews emphasize how untapped genetic variation in wild *Gossypium* species — including salt and wilt tolerance — can broaden the genetic base and enhance stress resilience in elite cultivars [41].

The combination of high-density SNP arrays, GWAS, and wild species introgression thus forms a comprehensive, modern toolkit for cotton breeding. These synergistic approaches accelerate the development of improved germplasm exhibiting superior agronomic traits and resilience to environmental stresses.

Conclusion. In this article, we first reviewed the genome assemblies of several key *G. hirsutum* cultivars, including TM-1, CRI-12, and ZM24, which have served as reference materials for comparative genomic analyses. We highlighted the pioneering works of Li et al., Zhang et al., and Wang et al., who provided high-quality assemblies that have significantly advanced our understanding of cotton genome organization, gene content, and structural variation. These genomic resources form the foundation for dissecting complex traits such as fiber quality, yield potential, and stress tolerance.

Furthermore, we examined recent advances in genomic tools and breeding strategies for *G. hirsutum* in greater detail. Particular attention was given to the integration of GWAS, high-density SNP arrays, and transcriptomic profiling which collectively enhance the resolution of trait mapping and accelerate the development of elite cultivars. The works of Fang et al., Huang et al., and Chen et al. exemplify the application of these approaches in identifying candidate genes linked to agronomically important traits, including salinity tolerance, drought resistance, and improved fiber properties.

By synthesizing current knowledge on genome assemblies and advanced breeding methodologies, this review underscores the critical importance of integrating genomic data into breeding programs. Such integration not only facilitates a more precise identification of key genetic determinants but also enables the design of molecular breeding strategies tailored to global agricultural challenges. Ultimately, these insights provide a robust framework for the sustainable improvement of *G. hirsutum* cultivars to meet the demands of both producers and the textile industry in the context of climate change and resource limitations.

Conflict of interest. The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Financing. The research was carried out under the Program-targeted financing of the Ministry of Science and Higher Education of the Republic of Kazakhstan for 2024-2026, AP23489921 «Creation of drought tolerance classification for Kazakhstan cotton-collection and identification of SNP markers associated with drought tolerance traits». Acknowledgments: We thank the Ministry of Science and Higher Education of the Republic of Kazakhstan for providing financial support.

References:

[1] **Zhang S.**, Cai Y., Guo J., Li K., Peng R., Liu F., Roberts J.A., Miao Y., Zhang X. Genotyping-by-sequencing of *Gossypium hirsutum* races and cultivars uncovers novel patterns of genetic relationships and domestication footprints // *Evol. Bioinform. Online*. – 2019. – Vol. 15. – Art. 1176934319889948. https://doi.org/10.1177/1176934319889948

[2] **Li F**., Fan G., Lu C., et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution // *Nat. Biotechnol.* – 2015. – Vol. 33. – P. 524–530. https://doi.org/10.1038/nbt.3208.

[3] **Zhang T.**, Hu Y., Jiang W., et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement // *Nat. Biotechnol.* – 2015. – Vol. 33. – P. 531–537. https://doi.org/10.1038/nbt.3207.

[4] **Sreedasyam A**., Lovell J.T., Mamidi S., et al. Genome resources for three modern cotton lines guide future breeding efforts // *Nat. Plants*. – 2024. – Vol. 10. – P. 1039–1051. https://doi.org/10.1038/s41477-024-01713-z.

[5] **Ai X., Liang Y.**, Wang J., Zheng J., Gong Z., Guo J., Li X., Qu Y. Genetic diversity and structure of elite cotton germplasm (*Gossypium hirsutum* L.) using genome-wide SNP data // *Genetica*. – 2017. – Vol. 145, №4–5. – P. 409–416. https://doi.org/10.1007/s10709-017-9976-8.

[6] **Chen Z.J.**, Sreedasyam A., Ando A., et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement // *Nat. Genet.* – 2020. – Vol. 52. – P. 525–533. https://doi.org/10.1038/s41588-020-0614-5.

[7] **Yang Z**., Ge X., Yang Z., et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars // *Nat. Commun.* – 2019. – Vol. 10. – Art. 2989. https://doi.org/10.1038/s41467-019-10820-x.

[8] **Saski C.A.**, Scheffler B.E., Hulse-Kemp A.M., et al. Subgenome anchored physical frameworks of the allotetraploid upland cotton (*Gossypium hirsutum* L.) genome, and an approach toward reference-grade assemblies of polyploids // *Sci. Rep.* – 2017. – Vol. 7, №1. – Art. 15274. https://doi.org/10.1038/s41598-017-14885-w.

[9] **Lu X.**, Wang X., Chen X., Shu N., Wang J., Wang D., Wang S., Fan W., Guo L., Guo X., Ye W. Single-base resolution methylomes of upland cotton (*Gossypium hirsutum* L.) reveal epigenome modifications in response to drought stress // *BMC Genomics*. – 2017. – Vol. 18, №1. – Art. 297. https://doi.org/10.1186/s12864-017-3681-y.

[10] **Gong Z.**, Xie F., Zhao T., Liang C., Zhang M., Wang J., Luo H., Wang L. Whole-genome bisulfite sequencing of upland cotton reveals drought-induced DNA methylation changes // *Genes*. – 2024. – Vol. 15, №6. – Art. 789. https://doi.org/10.3390/genes15060789.

[11] **Joshi B.**, Singh S., Tiwari G.J., et al. Genome-wide association study of fiber yield-related traits uncovers novel genomic regions and candidate genes in Indian upland cotton (*Gossypium hirsutum* L.) // *Front. Plant Sci.* – 2023. – Vol. 14. – Art. 1252746. https://doi.org/10.3389/fpls.2023.1252746.

[12] **Hou S.**, Yang L., Pan Y., et al. Genome-wide association studies reveal genetic variation underlying salt tolerance in upland cotton (*Gossypium hirsutum* L.) // *Front. Plant Sci.* – 2018. – Vol. 9. – Art. 1115. https://doi.org/10.3389/fpls.2018.01115.

[13] **Sun F.**, Zhang J., Li S., Wang T., Zhou Y., Liu X. Genome-wide association study of drought tolerance-related traits in upland cotton (*Gossypium hirsutum* L.) under field conditions // *BMC Genomics*. – 2023. – Vol. 24, №1. – Art. 645. https://doi.org/10.1186/s12864-023-09645-3.

[14] **Hu Y.**, Chen J., Fang L., et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton // *Nat. Genet.* – 2019. – Vol. 51, №4. – P. 739–748. https://doi.org/10.1038/s41588-019-0371-5.

[15] **Hulse-Kemp A.M**., Juenger T.E., Hazzouri K.M., et al. Development of a 63K SNP array for cotton and high-density mapping of intra- and interspecific populations of *Gossypium* spp. // *BMC Genomics*. – 2014. – Vol. 15. – Art. 2. https://doi.org/10.1186/1471-2164-15-2.

[16] **Du X.**, Huang G., He S., et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits // *Nat. Genet.* – 2018. – Vol. 50. – P. 796–802. https://doi.org/10.1038/s41588-018-0116-x.

[17] **Hou S.**, Zhu G., Li Y., Li W., Fu J., Niu E., et al. Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum* L.) // *Front. Plant Sci.* – 2018. – Vol. 9. – Art. 1276. https://doi.org/10.3389/fpls.2018.01276.

[18] **Li R.**, Ma X., Zhang Y., et al. Genome-wide identification and analysis of a cotton secretome reveals its role in resistance against *Verticillium dahliae* // *BMC Biology*. – 2023. – Vol. 21, No. 1. – Art. 166. https://doi.org/10.1186/s12915-023-01650-x.

[19] **Li Y**., Si Z., Wang G., et al. Genomic insights into the genetic basis of cotton breeding in China // *Molecular Plant*. – 2023. – Vol. 16. https://doi.org/10.1016/j.molp.2023.01.012.

[20] **Perkin L.C.**, Bell A., Hinze L.L., et al. Genome assembly of two nematode-resistant cotton lines (*Gossypium hirsutum* L.) // *G3: Genes|Genomes|Genetics*. – 2021. – Vol. 11, No. 11. – Art. jkab276. https://doi.org/10.1093/g3journal/jkab276.

[21] **Ma Z.**, Zhang Y., Wu L., et al. High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement // *Nature Genetics*. – 2021. – Vol. 53. – P. 1385–1391. https://doi.org/10.1038/s41588-021-00910-2.

[22] **Peng R.**, Xu Y., Tian S., et al. Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons // *Proceedings of the National Academy of Sciences of the USA*. – 2022. – Vol. 119, No. 39. – Art. e2208496119. https://doi.org/10.1073/pnas.2208496119.

[23] **Wang J.**, Liang Y., Gong Z., et al. Genomic and epigenomic insights into the mechanism of cold response in upland cotton (*Gossypium hirsutum*) // *Plant Physiology and Biochemistry*. – 2024. – Vol. 206. – Art. 108206. https://doi.org/10.1016/j.plaphy.2023.108206.

[24] **Xu Z.**, Wang G., Zhu X., et al. Genome assembly of two allotetraploid cotton germplasms reveals mechanisms of somatic embryogenesis and enables precise genome editing // *Nature Genetics*. – 2025. – Vol. 57. – P. 2028–2039. https://doi.org/10.1038/s41588-025-02258-3.

[25] **Sreedasyam A.**, Lovell J.T., Mamidi S., et al. Genome resources for three modern cotton lines guide future breeding efforts // *Nature Plants*. – 2024. – Vol. 10, No. 6. – P. 1039–1051. https://doi.org/10.1038/s41477-024-01713-z.

[26] **Cohen Z.**, Perkin L., Wagner T., et al. Nematode-resistance loci in upland cotton genomes are associated with structural differences // *Research Square*. – 2023. – 03 October. PREPRINT. Version 1. https://doi.org/10.21203/rs.3.rs-3377976/v1.

[27] **Bourland F.M.** Summaries of Arkansas Cotton Research 2024 // *Arkansas Agricultural Experiment Station Research Series*. – 2025. – URL: https://scholarworks.uark.edu/aaesser/235 (дата обращения: 05.09.2025).

[28] **Zhang H.**, Zhang W., Tang Y., et al. Mining key drought-resistant genes of upland cotton based on RNA-Seq and WGCNA analysis // *Plants*. – 2025. – Vol. 14, No. 10. – Art. 1407. https://doi.org/10.3390/plants14101407.

[29] **Sun F.**, Ma Y., Shi W., Yang Y. Genome-wide association analysis revealed genetic variation and candidate genes associated with the yield traits of upland cotton under drought conditions // *BMC Genomics*. – 2023. – Vol. 24. – Art. 531. https://doi.org/10.1186/s12864-023-09640-7.

[30] **Hinze L.L.**, Hulse-Kemp A.M., Wilson I.W., et al. Diversity analysis of cotton (*Gossypium hirsutum* L.) germplasm using the CottonSNP63K Array // *BMC Plant Biology*. – 2017. – Vol. 17, No. 1. – Art. 37. https://doi.org/10.1186/s12870-017-0981-y.

[31] **Tan Z.**, Zhang Z., Sun X., et al. Genetic map construction and fiber quality QTL mapping using the CottonSNP80K Array in upland cotton // *Frontiers in Plant Science*. – 2018. – Vol. 9. – Art. 225. https://doi.org/10.3389/fpls.2018.00225.

[32] **Chen Y.**, Gao Y., Chen P., et al. Genome-wide association study reveals novel quantitative trait loci and candidate genes of lint percentage in upland cotton based on the CottonSNP80K array // *Theoretical and Applied Genetics*. – 2022. – Vol. 135, No. 7. – P. 2279–2295. https://doi.org/10.1007/s00122-022-04111-1.

[33] **Huang C.**, Nie X., Shen C., et al. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs // *Plant Biotechnology Journal*. – 2017. – Vol. 15. – P. 1374–1386. https://doi.org/10.1111/pbi.12722.

[34] **Liang C.**, Meng Z., Meng Z., et al. GhABF2, a bZIP transcription factor, confers drought and salinity tolerance in cotton (*Gossypium hirsutum* L.) // *Scientific Reports*. – 2016. – Vol. 6. – Art. 35040. https://doi.org/10.1038/srep35040

[35] Thermo Fisher Scientific. Axiom Cotton Genotyping Array // URL: https://assets.thermofisher.com/TFSAssets/LSG/brochures/axiom_cotton_genotyping_array_datasheet.pdf (дата обращения: 05.09.2025).

[36] **Liang C.**, Liu Y., Li Y., et al. Activation of ABA receptor gene GhPYL9-11A is positively correlated with cotton drought tolerance in transgenic *Arabidopsis* // *Frontiers in Plant Science*. – 2017. – Vol. 8. – Art. 1453. https://doi.org/10.3389/fpls.2017.01453

[37] **Lin L.**, Pierce G.J., Bowers J.E., et al. A draft physical map of a D-genome cotton species (*Gossypium raimondii*) // *BMC Genomics*. – 2010. – Vol. 11. – Art. 395. https://doi.org/10.1186/1471-2164-11-395

[38] **Hou S**., Zhu G., Li Y., et al. Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum* L.) // *Frontiers in Plant Science*. – 2018. – Vol. 9. – Art. 1276. https://doi.org/10.3389/fpls.2018.01276

[39] **Guo M**., Zhang Y., Jia X., et al. Alternative splicing of REGULATOR OF LEAF INCLINATION 1 modulates phosphate starvation signaling and growth in plants // *Plant Cell*. – 2022. – Vol. 34, No. 9. – P. 3319–3338. https://doi.org/10.1093/plcell/koac161

[40] **Sun F.**, Ma Y., Shi W., Yang Y. Genome-wide association analysis revealed genetic variation and candidate genes associated with the yield traits of upland cotton under drought conditions // *BMC Genomics*. – 2023. – Vol. 24. – Art. 531. https://doi.org/10.1186/s12864-023-09640-7

[41] **Silva C.A.D**., Barroso P.A.V., Hoffmann L.V. Introgression of fiber quality traits from *Gossypium mustelinum* into upland cotton (*G. hirsutum* L.) by backcrossing // *Euphytica*. – 2017. – Vol. 213, No. 4. – Art. 85. https://doi.org/10.1007/s10681-017-1885-7

МАҚТА (*GOSSYPIUM HIRSUTUM*) ГЕНОМЫНЫҢ РЕСУРСТАРЫ ЗАМАНАУИ СЕЛЕКЦИЯЛЫҚ ЖҰМЫСТАР ҮШІН МОЛЕКУЛАЛЫҚ НЕГІЗ РЕТІНДЕ

Амангельдыева Н.К.[3], 7М05107 – «Биология» ББББ-ның 2 курс магистранты
Өркен А.[1,3], 8D05107 – «Биология» ББББ-ның 1-курс докторанты
Жумабай Н. Б.[1], 8D08100 – «Агрономия» ББББ-ның 1-курс докторанты
Манабаева Ш. А.[1,3], биология ғылымдарының кандидаты, профессор
Махмаджанов С. П.[2], ауылшаруашылық ғылымдарының кандидаты, қауымдастырылған профессор
Түсіпқан Д.[1*], PhD, қауымдастырылған профессор

[1]Ұлттық биотехнология орталығы, Астана қ., Қазақстан
[2]«Мақта және бақша ауыл шаруашылығы тәжірибе станциясы» ЖШС, Түркістан обл., Атакент а., Қазақстан
[3]Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

Аңдатпа. Соңғы он жылда геномды секвенирлеудегі жетістіктер *Gossypium hirsutum*-ды зерттеуді түбегейлі өзгертті, өйткені бұрынғы фрагменттелген қысқа жинақтар енді толық дерлік, хромосома деңгейіндегі геномдық карталарға айналды. Бұл шолуда *G. hirsutum*-ның негізгі культиварларының геномдық жинақтары қарастырылды, олардың ішінде алғашқы жоғары сапалы геномдық жинақты қамтамасыз еткен TM-1 эталондық линиясы негізгі орын алады. Бұдан кейін геном құрылымын, гендік құрамын және құрылымдық вариацияларды түсінуді жетілдірген CRI-12, ZM24 және басқа да элиталық селекциялық тектармақтары қарастырылды. Бұл геномдық ресурстар талшық сапасы, өнімділік және стресс төзімділігі сияқты маңызды белгілердің генетикалық негізін анықтауға мүмкіндік береді. Кейін геномдық құралдар мен селекция стратегияларындағы соңғы жетістіктері талқыланды. Әсіресе, GWAS зерттеулеріне, жоғары тығыздықтағы SNP массивтеріне, транскриптомдық профильдеуге ерекше көңіл бөлінді, себебі олар белгілерді карталауды және кандидат-гендерді анықтауды жеделдетеді. Геномдық жинақтар мен молекулалық селекциядағы прогресті қорытындылай отырып, бұл шолу тұрақты әрі жоғары өнімді сорттарды дамытуда геномдық ресурстардың негізгі рөлін айқындайды.

Тірек сөздер: Мақта (*Gossypium hirsutum*), толық геномды секвенирлеу, геномдық жинақ, пангеном, GWAS, молекулалық селекция, құрылымдық вариациялар.

ГЕНОМНЫЕ РЕСУРСЫ ХЛОПЧАТНИКА (*GOSSYPIUM HIRSUTUM*) КАК МОЛЕКУЛЯРНАЯ ОСНОВА ДЛЯ СОВРЕМЕННЫХ СЕЛЕКЦИОННЫХ РАЗРАБОТОК

Амангельдыева Н.К.[3], магистрант 2 курса ОП 7М05107 – «Биология»
Өркен А.[1,3], магистр естественных наук, докторант 1 курса ОП 8D05107 – «Биология»
Жумабай Н. Б.[1], магистр естественных наук, докторант 1 курса ОП 8D08100 – «Агрономия»
Манабаева Ш. А.[1,3], кандидат, биологических наук профессор
Махмаджанов С.П.[2], кандидат сельскохозяйственных наук, ассоциированный профессор
Түсіпқан Д.[1*], PhD, ассоциированный профессор

[1]*Национальный центр биотехнологии, Астана, Казахстан*
[2]*Сельскохозяйственная опытная станция хлопководства и бахчеводства, с Атакент, Туркестанская обл., Казахстан*
[3]*Евразийский национальный университет им. Л.Н. Гумилева, г. Астана, Казахстан*

Аннотация. Достижения в области секвенирования генома за последнее десятилетие радикально изменили наше понимание *Gossypium hirsutum*, превратив ранее фрагментированные короткочитные сборки в практически полные, хромосомные карты генома. В данном обзоре рассмотрены геномные сборки ключевых культиваров *G. hirsutum*, начиная с эталонной линии TM-1, которая послужила основой для первых высококачественных геномных сборок. Далее проанализированы такие культивары, как CRI-12, ZM24 и ряд элитных селекционных линий, чьи сборки уточнили знания о структуре генома, составе генов и структурных вариациях. Эти геномные ресурсы предоставляют важные сведения о генетической основе таких признаков, как качество волокна, урожайность и устойчивость к стрессам. Далее в обзоре обсуждаются современные геномные инструменты и стратегии селекции. Особое внимание уделено GWAS-исследованиям, высокоплотным SNP-массивам, транскриптомному профилированию, которые ускоряют картирование признаков и выявление кандидатных генов. Обобщая прогресс в области геномных сборок и молекулярной селекции, данный обзор подчеркивает центральную роль геномных ресурсов в создании устойчивых и высокоурожайных сортов, отвечающих мировым аграрным и промышленным вызовам.

Ключевые слова: Хлопчатник (*Gossypium hirsutum*), полное секвенирование генома, геномная сборка, пангеном, GWAS, молекулярная селекция, структурные вариации.