

## Review Article

## Open Access

# Bioinformatics Tools for Cotton Genomics: A Review

Zhen Li ✉

Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China

✉ Corresponding email: [zhen.li@hibio.org](mailto:zhen.li@hibio.org)Bioscience Evidence, 2025, Vol.15, No.5 doi: [10.5376/be.2025.15.0022](https://doi.org/10.5376/be.2025.15.0022)

Received: 24 Jul., 2025

Accepted: 29 Aug., 2025

Published: 15 Sep., 2025

**Copyright** © 2025 Li, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**Li Z., 2025, Bioinformatics tools for cotton genomics: a review, Bioscience Evidence, 15(5): 219-227 (doi: [10.5376/be.2025.15.0022](https://doi.org/10.5376/be.2025.15.0022))

**Abstract** The development of cotton genomics is closely related to new technologies. High-throughput sequencing enables us to obtain more data, and there are also an increasing number of bioinformatics tools. Some commonly used platforms nowadays include CottonFGD, CottonMD, CottonGen and CottonGVD. These databases collect information at different levels such as the genome, transcriptome, epigenome, metabolome and phenotype. Researchers can use them for gene function annotation, variation detection, trait analysis and molecular breeding. These tools have significantly enhanced research efficiency, helped us better understand complex traits, and promoted precise breeding. In the future, long-read sequencing, pan-genomics, multi-omics integration, and artificial intelligence will all bring new impetus to research. The aim of this study is to summarize the application of these tools and explore their role in precise cotton breeding.

**Keywords** Cotton genomics; Bioinformatics tools; Multi-omics database; Precision breeding; Mining of functional genes

## 1 Introduction

Cotton (*Gossypium* spp.) is one of the most important natural fiber crops in the world and also an important source of oil. It not only supports the textile industry, but also provides food and feed by-products (Yu et al., 2013; Yang et al., 2020). With the continuous increase in population and the pressure of climate change, the demand for high-yield, high-quality and stress-resistant cotton is growing. The development of genomics has provided new methods and tools for genetic improvement and trait research of cotton (Yu et al., 2013; Naoumkina and Kim, 2023). Through in-depth studies of whole-genome, transcriptomic and epigenomic data, scientists can identify genes related to important traits more quickly, promoting molecular breeding and precise selection (Yang et al., 2022; Khalilisamani et al., 2024).

However, the cotton genome is large and complex, and the amount of data is also very large. The traditional analytical methods are no longer sufficient. Bioinformatics tools have thus become key assistants in research. Nowadays, there are many databases and platforms, such as CottonGen, CottonFGD, CottonMD and CottonGVD. These platforms integrate multiple types of data such as genomics, transcriptomics, variations, epigenetics and metabolomics, and are also equipped with functions such as BLAST, gene annotation, QTL mapping, GWAS and co-expression networks (Yu et al., 2013; 2015; Zhu et al., 2017; Yang et al., 2020; Peng et al., 2021; Yang et al., 2022). These tools can conduct single-gene research, multi-omics and cross-species batch analysis, and visualize the results to assist in cotton functional genomics and molecular breeding research.

This article mainly sorts out the commonly used bioinformatics tools and applications in the current field of cotton genomics, with a focus on the integration and analysis of genomic, transcriptomic, epigenomic and variant data. We will also discuss the role of these tools in trait research, stress resistance exploration and molecular breeding, as well as the challenges faced in the research, and look forward to future development directions. It is hoped that this can provide references for researchers and promote the genetic improvement and industrial upgrading of cotton.

## 2 Cotton Genomics Landscape

### 2.1 Overview of cotton genome structure and evolution (diploid vs. allotetraploid)

The genus *Gossypium* has approximately 50 species. They were divided into eight diploid groups (A-G, K) and one allotetraploid group (AD) (Chen et al., 2017). The genomic structure of diploid cotton (such as *G. arboreum*

and *G. raimondii*) is relatively simple. Allotetraploid cotton (such as *G. hirsutum*, *G. barbadense*) is formed by the combination of two subgenomes, A and D, with doubled chromosome numbers and a more complex genome (Wang et al., 2018; Hu et al., 2019; Huang et al., 2020) (Figure 1). Tetraploid cotton was formed approximately 1 to 2 million years ago and has undergone genomic replication, chromosomal rearrangement, gene loss and transposition element amplification. These changes have led to significant differences among subgenomes (Pan et al., 2020). Among them, the amplification of transposition elements promotes genomic enlargement and also affects gene regulation and three-dimensional structure (Wang et al., 2021; 2022).

## 2.2 Key milestones in cotton genome sequencing projects

In 2012, the first cotton D genome (*G. raimondii*) was released. Subsequently, the reference genomes of genome A (*G. arboreum*, *G. herbaceum*) and allotetraploid cotton (*G. hirsutum*, *G. barbadense*) were also completed successively, greatly promoting the research (Hu et al., 2019; Huang et al., 2020). In recent years, new technologies such as single-molecule sequencing, Hi-C and optical mapping have made gene assembly more complete, especially in high-repetition regions and centromere regions (Wang et al., 2018; Hu et al., 2019; Pan et al., 2020). Meanwhile, whole-genome resequencing, pan-genome and three-dimensional genome studies have revealed the diversity and regulatory networks of the cotton genome (Pan et al., 2020; Wang et al., 2022; Li et al., 2024).

## 2.3 Public databases and resources (CottonGen, CottonFGD, NCBI, Ensembl Plants, etc.)

Cotton genome research relies on multiple databases and platforms (Pan et al., 2020):

CottonGen: Integrated genomic, transcriptomic, variant, QTL and phenotypic data, providing multiple analytical tools.

CottonFGD: Mainly used for functional genome annotation, including gene families, expression profiles and visualization tools.

NCBI, Ensembl Plants: Provide genomic sequence, annotation and alignment tools for cotton and related species.

These databases facilitate data storage, retrieval and sharing, enabling researchers to use genomic information more efficiently.

## 2.4 Challenges specific to cotton genomics (polyploidy, repetitive sequences).

The cotton genome is highly polyploid and contains a large number of repetitive sequences. This brings great difficulty to assembly, annotation and mutation detection (Wang et al., 2018; Hu et al., 2019; Huang et al., 2020; Pan et al., 2020). The A and D subgenomes of allotetraploids have many homologous genes and structural variations, which can easily cause confusion in gene attribution and make expression regulation more complex. Furthermore, transposition elements occupy the majority of the genome (more than 70% of the A genome), increasing instability and making annotation more difficult (Pan et al., 2020; Wang et al., 2021). Therefore, more efficient and precise bioinformatics tools are needed to enhance data parsing capabilities.

# 3 Categories of Bioinformatics Tools

## 3.1 Genome assembly and annotation

High-quality assembly and annotation are the foundation for studying the cotton genome. Commonly used sequence alignment software includes Bowtie2 and BWA, and there are also some automated annotation platforms. CottonGen, CottonFGD and CottonMD integrate the reference genomes and annotation data of multiple cotton species and can be used for gene mapping, structural analysis and batch annotation (Yu et al., 2013; Zhu et al., 2017; Yang et al., 2020; 2022).

## 3.2 Comparative genomics and evolutionary analysis

Comparative genomic tools (such as MCScanX, VISTA, and PLAZA) are often used to analyze homology and collinearity among different species or subgenomes, and can also be used to study evolutionary relationships. These analyses revealed gene family expansion, genome-wide replication and structural variations (Gao et al., 2017; Yang et al., 2020; Su et al., 2023). The CottonGen et al. Platform provides genomic alignment and genetic map comparison, facilitating the study of evolution and polyploidy (Yu et al., 2013; 2015; Yang et al., 2020).

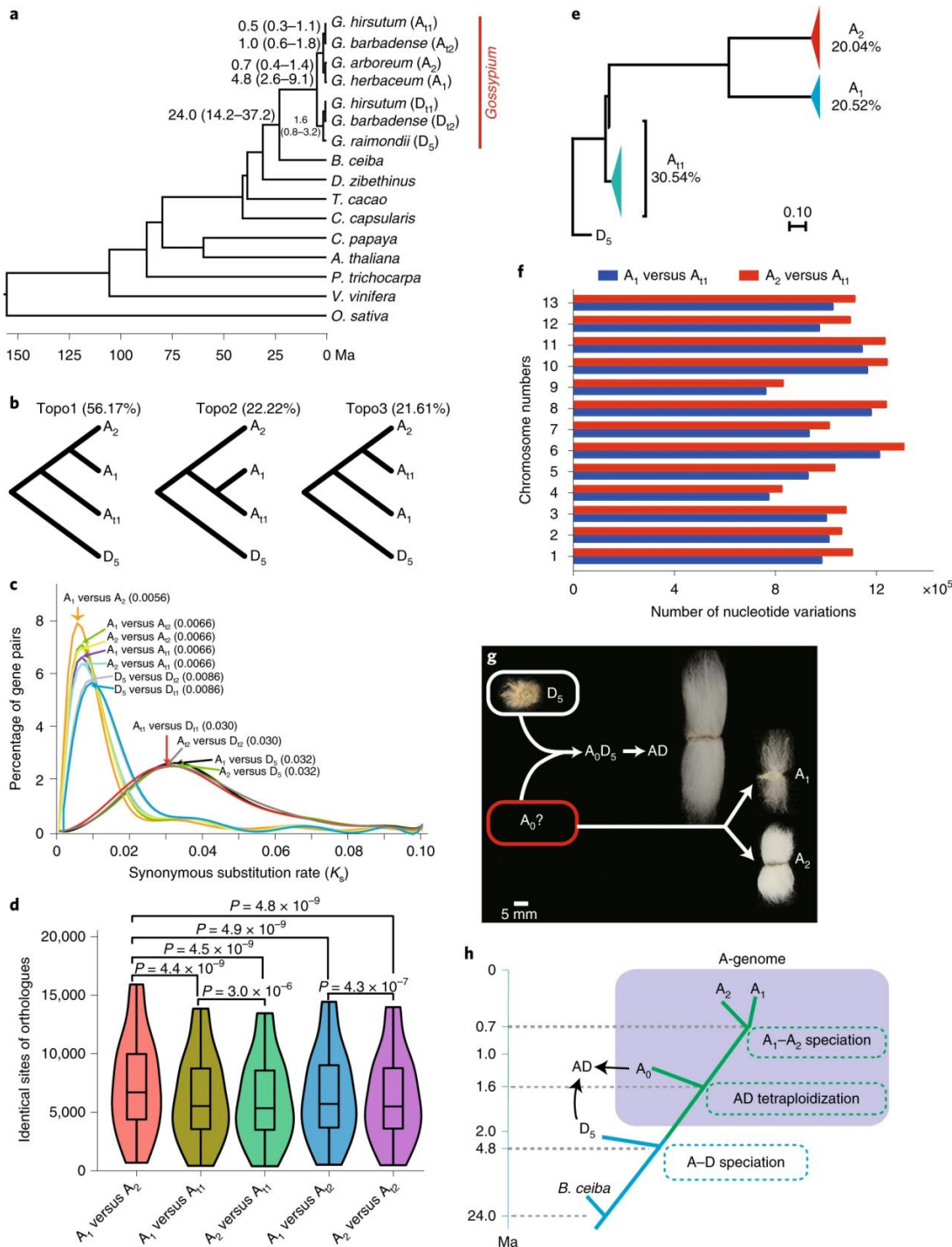


Figure 1 The evolution of the allotetraploid cotton genome (Adopted from Huang et al., 2020)

Image caption: a, Inferred phylogenetic analysis among *Gossypium* and other eudicot plants. b, Summary of phylogenetic analysis with the approximately unbiased test in 10-kb windows. c, Distribution of  $K_s$  values for orthologous genes among cotton genomes. Peak values for each comparison are indicated in the parentheses. d, Comparisons of identical sites in orthologous genes. Violin plots summarize the distribution of identical sites. The center line in each box indicates the median, and the box limits indicate the upper and lower quartiles of divergence ( $n=20$  types of synonymous mutation). P values were derived with Student's t-test. e, Phylogenetic and ancestral allele analysis based on SNPs. The red, blue and green triangles represent the collapsed 21  $A_2$  accessions, 14  $A_1$  accessions and 30 (AD)1 accessions, respectively. The percentage value indicates the percentage of ancestral alleles for each species that were identical to those of the  $D_5$ -genome. f, Number of nucleotide variations in  $A_1$  or  $A_2$  compared with  $A_{11}$  across the chromosomes. g, A model for the formation of allotetraploid cotton showing fiber phenotypes from the (AD)1 (accession TM-1), the  $D_5$ , the  $A_1$  (var. africanum) and the  $A_2$  (cv. Shixiya1). Scale bar, 5 mm. h, A schematic map of the evolution of cotton genomes. Major evolutionary events are shown in dashed boxes (Adopted from Huang et al., 2020)

### **3.3 Transcriptomics and gene expression profiling**

Transcriptomic methods (such as RNA-Seq process, DESeq2 and co-expression network analysis) are widely used in the expression study of cotton. The research subjects include gene expression under different tissues, developmental stages and stress conditions. CottonFGD and CottonMD integrate a large amount of expression data and provide visualization tools, which can help study expression patterns and functional annotations (Zhu et al., 2017; Yang et al., 2020; 2022; Khalilisamani et al., 2024).

### **3.4 Variant detection and genotyping**

The detection and typing of variations such as SNPS and Indels are the focus of molecular breeding and population genetics research. CottonGVD and CottonMD integrate large-scale resequencing data and are equipped with built-in tools such as GWAS, eGWAS, and SNPmatch. Researchers can use them for variant screening, trait association and population structure analysis (Yu et al., 2015; Peng et al., 2021; Yang et al., 2022).

### **3.5 Epigenomics and regulatory genomics**

Epigenomic methods (such as DNA methylation, histone modification and chromatin accessibility analysis) have gradually been applied to cotton research. They help reveal the mechanisms of gene regulation and adaptation to adversity. CottonMD includes various epigenome data and provides visualization and correlation tools (Yang et al., 2020; 2022).

### **3.6 Systems biology and network analysis**

Systems biology methods (such as co-expression networks, metabolic pathways, and functional enrichment analysis) can integrate multi-omics data to study the molecular mechanisms of complex traits. CottonFGD and CottonMD provide network analyses at the levels of gene families, pathways, QTL and GWAS (Zhu et al., 2017; Yang et al., 2020; 2022; Khalilisamani et al., 2024).

### **3.7 Machine learning and AI in cotton genomics**

In recent years, machine learning and AI methods have begun to be applied in cotton genomics. They can be used for gene function prediction, phenotype prediction and precision breeding. For example, gene network clustering or Bayesian models based on transcriptome and SNP data can improve the prediction accuracy of complex traits and also provide new ideas for molecular design breeding (Yang et al., 2020; Khalilisamani et al., 2024).

## **4 Case Study: Application of Bioinformatics Tools in Cotton Research**

### **4.1 Topic example: genome-wide association study (GWAS) for fiber quality traits**

Genome-wide association analysis (GWAS) is an important method for studying the complex traits of cotton. These properties include fiber length and strength, etc. CottonMD is a commonly used platform that has collected genotype, phenotype, transcriptome and epigenome data of 4,180 cotton materials. The platform is equipped with a variety of statistical tools and can conduct analyses such as GWAS and eGWAS. Researchers can use these tools to quickly identify variations related to fiber quality and locate candidate genes. Meanwhile, more in-depth research can be conducted by combining expression data and functional annotations, thereby accelerating molecular breeding of superior fiber traits (Yang et al., 2022). Platforms such as CottonGen also provide result visualization, QTL localization and genome browsing functions, which are very convenient to use (Yu et al., 2013; 2015).

### **4.2 Alternative case study: Integrating transcriptomics and epigenomics to understand cotton stress response**

The study of how cotton responds to adverse conditions such as drought, salinity and low temperatures is increasingly relying on multi-omics integration. CottonMD integrated transcriptome data from 76 tissues, epigenome data (DNA methylation, histone modification, chromatin accessibility) from 5 species, and metabolome data. These resources can support the combined analysis of gene expression, regulatory elements and phenotypes. Researchers can leverage this platform to analyze the changes in gene expression and epigenetic modifications under stress conditions, and identify key regulatory networks and responding genes. These results provide a basis for the breeding of cotton varieties with high stress resistance (Yang et al., 2022). In addition,

some studies have also used machine learning methods to predict cold stress-related genes, improving the efficiency and accuracy of candidate gene screening (Tian et al., 2023).

## 5 Challenges and Limitations

### 5.1 Data complexity due to polyploidy and large genome size

The cotton genome is large, especially the allotetraploid *G. hirsutum* and *G. barbadense*, with a size of approximately 2.5 Gb (Yang et al., 2022; Manivannan and Amal, 2023; Kumar et al., 2024; Sheri et al., 2025). They have many repetitive genes and homologous genes, which makes assembly, annotation and functional research all more difficult. Polyploidy also makes the regulation of gene expression more complex and increases the uncertainty in variant detection and gene editing. In addition, the proportion of repetitive sequences and transposic elements in the genome is very high, which also makes data analysis more difficult.

### 5.2 Need for cotton-specific annotation resources

Although there are databases such as CottonGen and CottonFGD, the functional annotation, phenotypic association and multi-omics integration resources for cotton are not yet complete (Ashraf et al., 2018; Li et al., 2021). The genomic versions published by different research groups vary in chromosome length and gene annotation, which can affect the accuracy of data alignment and analysis (Ashraf et al., 2018). Therefore, higher-quality and standardized annotation resources are needed, and a multi-species integration platform should also be established to better support gene mining and molecular breeding.

### 5.3 Limitations in computational power and reproducibility

Cotton genomics research requires handling large-scale multi-omics data, including genomic, transcriptomic, epigenomic and variomic data, etc. This poses very high requirements for computing power and algorithm efficiency (Yang et al., 2022; Manivannan and Amal, 2023). The current process often encounters problems such as insufficient memory, tight storage and slow speed when dealing with extremely large data. Furthermore, some processes and databases lack unified standards and version management, and the results are not easy to reproduce and compare (Ashraf et al., 2018; Yang et al., 2022). Therefore, it is necessary to establish more efficient, scalable and standardized processes to enhance research efficiency and reliability.

### 5.4 Data sharing and FAIR principles in cotton genomics

With the rapid increase of multi-omics data, data sharing and repeatability have become the focus of international attention. The application of the FAIR principle (searchable, accessible, interoperable, reusable) in cotton genomics is not sufficient (Ashraf et al., 2018; Li et al., 2021). Some data and results were not made public in a timely manner, or there was a lack of a unified metadata standard, which all affected the circulation and reuse of data. Promoting data standardization, open sharing and cross-platform interoperability is the key to enhancing international cooperation and research innovation.

## 6 Future Directions

### 6.1 Advances in long-read sequencing and pan-genomics for cotton

The development of long-read sequencing (LRS) has greatly promoted the research of cotton genomics. LRS can generate very long sequence reads, making gene assembly more continuous and accurate, especially suitable for plant genomes with a large number of repetitive sequences. With the improvement of sequencing accuracy and throughput, LRS has become an important method for detecting structural variations, whole-genome assembly and pan-genome research. It provides new ideas for us to understand the germplasm diversity and functional gene variations of cotton (Amarasinghe et al., 2020; Logsdon et al., 2020; Amarasinghe et al., 2021; De Coster et al., 2021; Van Dijk et al., 2023). Pan-genomics, combined with multiple genomic data, can comprehensively demonstrate the genetic diversity within cotton species and lay the foundation for molecular breeding and functional gene research (Mascher et al., 2021).

### 6.2 Cloud-based and user-friendly bioinformatics platforms

With the rapid growth of multi-omics data, cloud computing and one-stop analysis platforms are becoming increasingly important. Cloud platforms such as Majorbio Cloud and Galaxy integrate multi-omics analysis processes, visualization tools and online learning modules, lowering the threshold of data analysis and facilitating



the processing and sharing of large-scale data (Liu et al., 2014; de Koning et al., 2020; Oh et al., 2020; Ren et al., 2022). These platforms not only enable researchers without a computing background to use them, but also promote data standardization and repeatability, which is of great help to the popularization and cooperation of cotton genomics.

### 6.3 Integration of multi-omics and machine learning for predictive breeding

If multi-omics data (genomic, transcriptomic, epigenomic, metabolomic, etc.) are combined with machine learning and deep learning, they can be better applied to precision breeding and trait prediction. Machine learning can be used for dimensionality reduction, regulatory network construction, candidate gene screening, and can also help establish an association model of genotype - phenotype - environment. These methods provide tools for molecular design breeding of complex traits (Oh et al., 2020; Yang et al., 2021; 2022; Yan and Wang, 2022). Platforms like CottonMD have integrated multi-omics data and analysis tools, providing significant support for cotton genetic improvement and functional gene research (Yang et al., 2022).

### 6.4 Need for international collaboration and open-source tool development

As the volume of data and the difficulty of analysis increase, international cooperation and the development of open-source tools become even more important. Open databases and tools (such as long-read-tools.org) facilitate resource sharing, method standardization and community communication (Amarasinghe et al., 2020; Amarasinghe et al., 2021; Thriveni et al., 2024) (Figure 2). Global cooperation helps to enhance data interoperability, promote the development of new algorithms and platforms, and accelerate the application and innovation of cotton genomics worldwide (Thriveni et al., 2024).

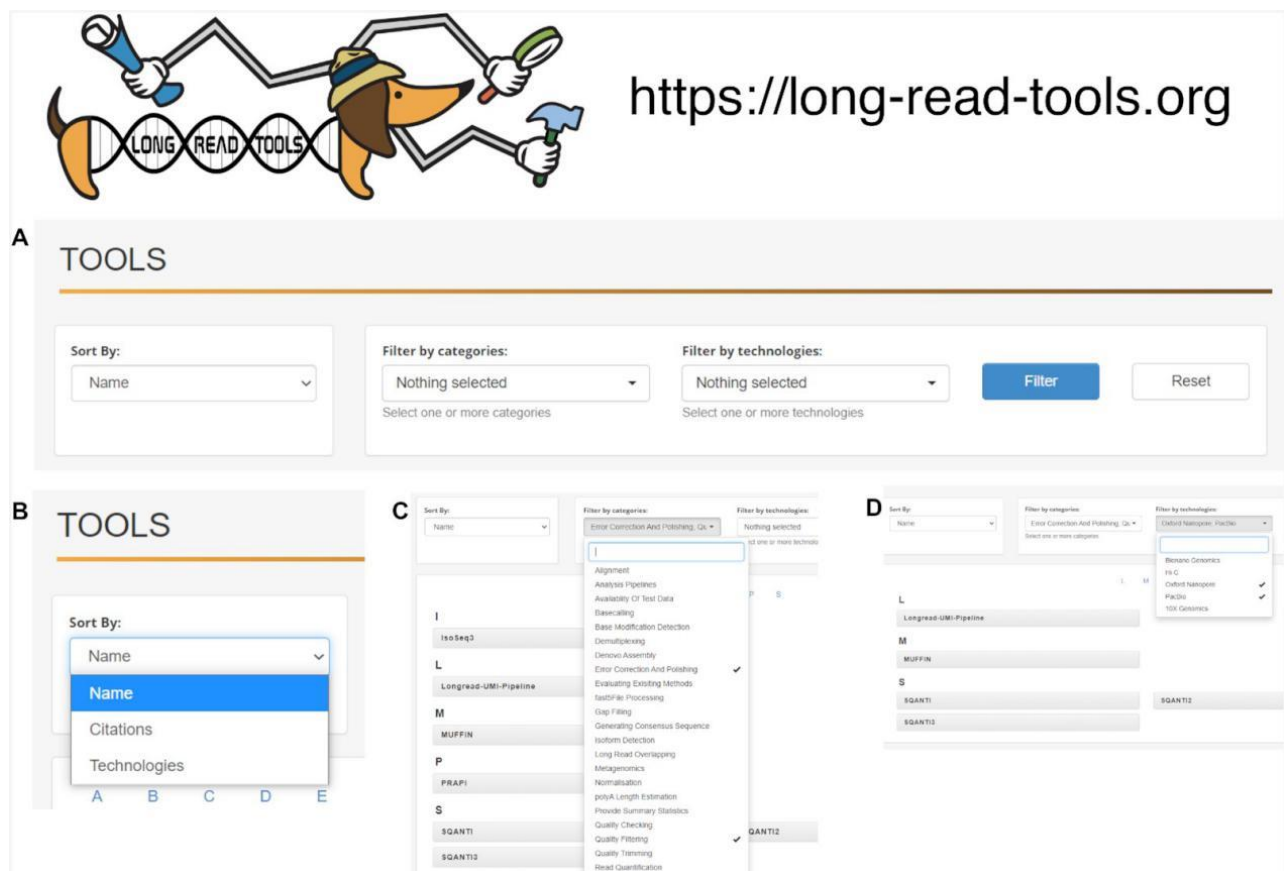


Figure 2 Example use of the Tools tab from long-read-tools.org. A. The custom toolbar for the page. B. Drop-down "Sort By" menu. C. Drop-down "Filter by categories" menu, which allows users to select multiple options by clicking on an item or typing the word in the text box. D. Drop-down Filter by "technologies" menu, which allows users to select multiple options by clicking on an item or typing the word in the text box. When multiple categories or technologies are selected, the website returns the intersection, not the union; i.e., a tool has to satisfy all the requirements to be reported (Adopted from Amarasinghe et al., 2021)

## 7 Conclusion

In recent years, many new tools have emerged in cotton genomics. Common ones include high-throughput sequencing platforms, genome assembly and annotation software, molecular marker development methods, genome-wide association studies (GWAS), as well as multi-omics databases (such as CottonGen, CottonFGD, COTTONOMICS) and phenotypic analysis systems based on deep learning. These tools have been widely applied in cotton research, such as genome sequencing, gene function annotation, trait mapping, mining of stress-resistant genes, molecular breeding and high-throughput phenotypic analysis, etc. They make genetic improvement faster and more precise.

With the development of bioinformatics, cotton research is shifting from traditional phenotypic selection to genome-driven precision breeding. Genomic sequencing and big data analysis have revealed the complex polyploid genomic structure of cotton and also demonstrated the relationship between functional gene regulatory networks and traits. These achievements have provided new ideas for high-yield, high-quality and stress-resistant breeding. The application of gene editing tools such as CRISPR/Cas enables researchers to precisely modify target genes, significantly accelerating the breeding of new varieties. Meanwhile, the application of deep learning and artificial intelligence in phenotypic identification, trait prediction and data mining has also promoted the automation and intelligence of research.

In the future, cotton genomics will still rely on these tools to further promote precise breeding and sustainable production. High-quality reference genomes and pan-genome resources have laid the foundation for trait improvement and diversity conservation. The combination of multi-omics integration, machine learning and big data analysis will help achieve molecular design breeding of complex traits. This not only increases output, but also improves quality and stress resistance. New technologies such as gene editing and molecular marker selection will also accelerate the breeding of disease-resistant and stress-resistant varieties, helping the cotton industry cope with climate change and resource pressure, and move towards green, efficient and sustainable development.

## Acknowledgments

I wish to express my sincere gratitude to the anonymous reviewers for their insightful feedback and constructive suggestions, which have greatly enhanced the quality of this paper.

## Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Amarasinghe S., Ritchie M., and Gouil Q., 2021, long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data, *GigaScience*, 10(2): giab003.  
<https://doi.org/10.1093/gigascience/giab003>
- Amarasinghe S., Su S., Dong X., Zappia L., Ritchie M., and Gouil Q., 2020, Opportunities and challenges in long-read sequencing data analysis, *Genome Biology*, 21: 30.  
<https://doi.org/10.1186/s13059-020-1935-5>
- Ashraf J., Zuo D., Wang Q., Malik W., Zhang Y., Abid M., Cheng H., Yang Q., and Song G., 2018, Recent insights into cotton functional genomics: progress and future perspectives, *Plant Biotechnology Journal*, 16: 699-713.  
<https://doi.org/10.1111/pbi.12856>
- Chen Z., Grover C., Li P., Wang Y., Nie H., Zhao Y., Wang M., Liu F., Zhou Z., Wang X., Cai X., Wang K., Wendel J., and Hua J., 2017, Molecular evolution of the plastid genome during diversification of the cotton genus, *Molecular Phylogenetics and Evolution*, 112: 268-276.  
<https://doi.org/10.1016/j.ympev.2017.04.014>
- De Coster W., Weissensteiner M., and Sedlazeck F., 2021, Towards population-scale long-read sequencing, *Nature Reviews. Genetics*, 22: 572-587.  
<https://doi.org/10.1038/s41576-021-00367-3>
- De Koning W., Miladi M., Hiltmann S., Heikema A., Hays J., Flemming S., van den Beek M., Mustafa D., Backofen R., Grüning B., and Stubbs A., 2020, NanoGalaxy: Nanopore long-read sequencing data analysis in Galaxy, *GigaScience*, 9(10): giaa105.  
<https://doi.org/10.1093/gigascience/giaa105>
- Gao W., Long L., Tian X., Xu F., Liu J., Singh P., Botella J., and Song C., 2017 Genome editing in cotton with the CRISPR/Cas9 system, *Frontiers in Plant Science*, 8: 1364.  
<https://doi.org/10.3389/fpls.2017.01364>

- Hu Y., Chen J., Fang L., Zhang Z., Niu Y., Ju L., Deng J., Zhao T., Lian J., Baruch K., Fang D., Liu X., Ruan Y., Rahman M., Han J., Wang K., Wang Q., Wu H., Mei G., Zang Y., Han Z., Xu C., Shen W., Yang D., Si Z., Dai F., Zou L., Huang F., Bai Y., Zhang Y., Brodt A., Ben-Hamo H., Zhu X., Zhou B., Guan X., Zhu S., Chen X., and Zhang T., 2019, *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton, *Nature Genetics*, 51: 739-748.  
<https://doi.org/10.1038/s41588-019-0371-5>
- Huang G., Wu Z., Percy R., Bai M., Li Y., Frelichowski J., Hu J., Wang K., Yu J., and Zhu Y., 2020, Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution, *Nature Genetics*, 52: 516-524.  
<https://doi.org/10.1038/s41588-020-0607-4>
- Khalilismami N., Li Z., Pettolino F., Moncuquet P., Reverter A., and MacMillan C., 2024, Leveraging transcriptomics-based approaches to enhance genomic prediction: integrating SNPs and gene networks for cotton fibre quality improvement, *Frontiers in Plant Science*, 15: 1420837.  
<https://doi.org/10.3389/fpls.2024.1420837>
- Kumar R., Das J., Puttaswamy R., Kumar M., Balasubramani G., and Prasad Y., 2024, Targeted genome editing for cotton improvement: prospects and challenges, *The Nucleus*, 67: 181-203.  
<https://doi.org/10.1007/s13237-024-00479-1>
- Li J., Liu Z., You C., Qi Z., You J., Grover C., Long Y., Huang X., Lu S., Wang Y., Zhang S., Wang Y., Bai R., Zhang M., Jin S., Nie X., Wendel J., Zhang X., and Wang M., 2024, Convergence and divergence of diploid and tetraploid cotton genomes, *Nature Genetics*, 56: 2562-2573.  
<https://doi.org/10.1038/s41588-024-01964-8>
- Li J., Yuan D., Wang P., Wang Q., Sun M., Liu Z., Si H., Xu Z., Zhang B., Pei L., Tu L., Zhu L., Chen L., Lindsey K., Zhang X., Jin S., and Wang M., 2021, Cotton pan-genome retrieves the lost sequences and genes during domestication and selection, *Genome Biology*, 22: 119.  
<https://doi.org/10.1186/s13059-021-02351-w>
- Liu B., Madduri R., Sotomayor B., Chard K., Lacinski L., Davé U., Li J., Liu C., and Foster I., 2014, Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses, *Journal of Biomedical Informatics*, 49: 119-133.  
<https://doi.org/10.1016/j.jbi.2014.01.005>
- Logsdon G., Vollger M., and Eichler E., 2020, Long-read human genome sequencing and its applications, *Nature Reviews Genetics*, 21: 597-614.  
<https://doi.org/10.1038/s41576-020-0236-x>
- Manivannan A., and Amal T., 2023, Deciphering the complex cotton genome for improving fiber traits and abiotic stress resilience in sustainable agriculture, *Molecular Biology Reports*, 50: 6937-6953.  
<https://doi.org/10.1007/s11033-023-08565-4>
- Mascher M., Wicker T., Jenkins J., Plott C., Lux T., Koh C., Ens J., Gundlach H., Boston L., Tulpová Z., Holden S., Hernández-Pinzón I., Scholz U., Mayer K., Spannagl M., Pozniak C., Sharpe A., Šimková H., Moscou M., Grimwood J., Schmutz J., and Stein N., 2021, Long-read sequence assembly: a technical evaluation in barley, *The Plant Cell*, 33: 1888-1906.  
<https://doi.org/10.1093/plcell/koab077>
- Naoumkina M., and Kim H., 2023, Bridging molecular genetics and genomics for cotton fiber quality improvement, *Crop Science*, 63(4): 1794-1815.  
<https://doi.org/10.1002/csc2.20987>
- Oh M., Park S., Kim S., and Chae H., 2020, Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations, *Briefings in Bioinformatics*, 22(1): 66-76.  
<https://doi.org/10.1093/bib/bbaa032>
- Pan Y., Meng F., and Wang X., 2020, Sequencing multiple cotton genomes reveals complex structures and lays foundation for breeding, *Frontiers in Plant Science*, 11: 560096.  
<https://doi.org/10.3389/fpls.2020.560096>
- Peng Z., Li H., Sun G., Dai P., Geng X., Wang X., Zhang X., Wang Z., Jia Y., Pan Z., Chen B., Du X., and He S., 2021, CottonGVD: A Comprehensive genomic variation database for cultivated cottons, *Frontiers in Plant Science*, 12: 803736.  
<https://doi.org/10.3389/fpls.2021.803736>
- Ren Y., Yu G., Shi C., Liu L., Guo Q., Han C., Zhang D., Zhang L., Liu B., Gao H., Zeng J., Zhou Y., Qiu Y., Wei J., Luo Y., Zhu F., Li X., Wu Q., Li B., Fu W., Tong Y., Meng J., Fang Y., Dong J., Feng Y., Xie S., Yang Q., Yang H., Wang Y., Zhang J., Gu H., Xuan H., Zou G., Luo C., Huang L., Yang B., Dong Y., Zhao J., Han J., Zhang X., and Huang H., 2022, Majorbio Cloud: A one-stop, comprehensive bioinformatic platform for multiomics analyses, *iMeta*, 1(2): e12.  
<https://doi.org/10.1002/imt2.12>
- Sheri V., Mohan H., Jogam P., Alok A., Rohela G., and Zhang B., 2025, CRISPR/Cas genome editing for cotton precision breeding: mechanisms, advances, and prospects, *Journal of Cotton Research*, 8: 4.  
<https://doi.org/10.1186/s42397-024-00206-w>
- Su J., Song S., Wang Y., Zeng Y., Dong T., Ge X., and Duan H., 2023, Genome-wide identification and expression analysis of DREB family genes in cotton, *BMC Plant Biology*, 23: 169.  
<https://doi.org/10.1186/s12870-023-04180-4>
- Thriveni V., Teotia J., Hazra S., Bharti T., Kumar M., Lallawmkimi M., and Panwar D., 2024, A review on integrating bioinformatics tools in modern plant breeding, *Archives of Current Research International*, 24(9): 293-308.  
<https://doi.org/10.9734/acri/2024/v24i9894>



- Tian Y., Shuai Y., Shao C., Wu H., Fan L., Li Y., Chen X., Narimanov A., Usmanov R., and Baboeva S., 2023, Extraction of cotton information with optimized phenology-based features from Sentinel-2 images, *Remote. Sens.*, 15: 1988.  
<https://doi.org/10.3390/rs15081988>
- Van Dijk E., Naquin D., Gorrichon K., Jaszczyszyn Y., Ouazahrour R., Thermes C., and Hernandez C., 2023, Genomics in the long-read sequencing era, *Trends in genetics: TIG*, 39(9): 649-671.  
<https://doi.org/10.1016/j.tig.2023.04.006>
- Wang M., Li J., Qi Z., Long Y., Pei L., Huang X., Grover C., Du X., Xia C., Wang P., Liu Z., You J., Tian X., Wang R., Chen X., He X., Fang D., Sun Y., Tu L., Jin S., Zhu L., Wendel J., and Zhang X., 2022, Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*, *Nature Genetics*, 54: 1959-1971.  
<https://doi.org/10.1038/s41588-022-01237-2>
- Wang M., Li J., Wang P., Liu F., Liu Z., Zhao G., Xu Z., Pei L., Grover C., Wendel J., Wang K., and Zhang X., 2021, Comparative genome analyses highlight transposon-mediated genome expansion and the evolutionary architecture of 3D genomic folding in cotton, *Molecular Biology and Evolution*, 38: 3621-3636.  
<https://doi.org/10.1093/molbev/msab128>
- Wang M., Tu L., Yuan D., Zhu D., Shen C., Li J., Liu F., Pei L., Wang P., Zhao G., Ye Z., Huang H., Yan F., Zhang L., Liu M., You J., Yang Y., Liu Z., Huang F., Li B., Qiu P., Zhang Q., Zhu L., Jin S., Yang X., Min L., Li G., Chen L., Zheng H., Lindsey K., Lin Z., Udall J., and Zhang X., 2018, Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*, *Nature Genetics*, 51: 224-229.  
<https://doi.org/10.1038/s41588-018-0282-x>
- Yan J., and Wang X., 2022, Machine learning bridges omics sciences and plant breeding, *Trends in Plant Science*, 28(2): 199-210.  
<https://doi.org/10.1016/j.tplants.2022.08.018>
- Yang Y., Saand M., Huang L., Abdelaal W., Zhang J., Wu Y., Li J., Sirohi M., and Wang F., 2021, Applications of multi-omics technologies for crop improvement, *Frontiers in Plant Science*, 12: 563953.  
<https://doi.org/10.3389/fpls.2021.563953>
- Yang Z., Qanmber G., Wang Z., Yang Z., and Li F., 2020, *Gossypium* genomics: trends, scope, and utilization for cotton improvement, *Trends in Plant Science*, 25(5): 488-500.  
<https://doi.org/10.1016/j.tplants.2019.12.011>
- Yang Z., Wang J., Huang Y., Wang S., Wei L., Liu D., Weng Y., Xiang J., Zhu Q., Yang Z., Nie X., Yu Y., Yang Z., and Yang Q., 2022, CottonMD: a multi-omics database for cotton biological study, *Nucleic Acids Research*, 51: D1446-D1456.  
<https://doi.org/10.1093/nar/gkac863>
- Yu J., Jung S., Cheng C., Ficklin S., Lee T., Zheng P., Jones D., Percy R., and Main D., 2013, CottonGen: a genomics, genetics and breeding database for cotton research, *Nucleic Acids Research*, 42: D1229-D1236.  
<https://doi.org/10.1093/nar/gkt1064>
- Yu J., Jung S., Cheng C., Lee T., Zheng P., Buble K., Crabb J., Humann J., Hough H., Jones D., Campbell J., Udall J., and Main D., 2015, CottonGen: the community database for cotton genomics, genetics, and breeding research, *Plants*, 10(12): 2805.  
<https://doi.org/10.3390/plants10122805>
- Zhu T., Liang C., Meng Z., Sun G., Meng Z., Guo S., and Zhang R., 2017, CottonFGD: an integrated functional genomics database for cotton, *BMC Plant Biology*, 17: 101.  
<https://doi.org/10.1186/s12870-017-1039-x>



#### Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.