

Research Progress in Genome Sequencing and Functional Gene Mining of Cannabis

Shanyu Chen ^{1*}, Huijuan Tang ^{2*}, Si Jie ¹, Wenjun Wang ³, Lijuan Tang ³, Guanhai Ruan ¹ ✉

¹ Institute of Crops and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, 310021, China;

² Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, 410205, China;

³ Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin, 150000, China

* These authors contributed equally to this work

✉ Corresponding author: 13906520484@163.com

Genomics and Applied Biology, 2024, Vol.15, No.5 doi: [10.5376/gab.2024.15.0024](https://doi.org/10.5376/gab.2024.15.0024)

Received: 07 Jul., 2024

Accepted: 19 Aug., 2024

Published: 08 Sep., 2024

Copyright © 2024 Chen et al., 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:

Chen S.Y., Tang H.J., Jie S., Wang W.J., Tang L.J., and Ruan G.H., 2024, Research progress in genome sequencing and functional gene mining of cannabis, *Genomics and Applied Biology*, 15(5): 223-234 (doi: [10.5376/gab.2024.15.0024](https://doi.org/10.5376/gab.2024.15.0024))

Abstract The primary goal of this study is to advance the understanding of the *Cannabis sativa* genome and to identify functional genes that contribute to its medicinal, industrial, and agricultural applications. Our comprehensive analysis revealed several key findings. Current Cannabis genome assemblies are incomplete, with significant portions missing or unmapped, which hampers accurate gene annotation. Recent advancements in genomics have identified four genes significantly associated with lifetime cannabis use: NCAM1, CADM2, SCOC, and KCNT2, which are linked to various phenotypes such as substance use and body mass index. Additionally, a high-quality reference genome for wild *Cannabis sativa* has been developed, providing valuable genetic resources for future research. In silico approaches have been proposed for genome editing, targeting genes involved in cannabinoid biosynthesis, which could lead to novel applications in agriculture and medicine. Furthermore, virus-induced gene silencing (VIGS) methods have been successfully applied to study gene functions in cannabis, demonstrating the potential for functional gene studies. The findings underscore the importance of coordinated efforts to complete and refine Cannabis genome assemblies. The identification of key genes and the development of advanced genomics tools hold significant promise for the genetic improvement of cannabis. These advancements could lead to enhanced medicinal and industrial applications, ultimately benefiting various sectors including agriculture, pharmaceuticals, and biotechnology.

Keywords *Cannabis sativa*; Genome sequencing; Functional gene mining; Genomics, Cannabinoid biosynthesis; Gene editing; Virus-induced gene silencing

1 Introduction

Cannabis sativa L., commonly known as cannabis, is a versatile plant species with a rich history of use spanning recreational, medicinal, industrial, and agricultural domains. It belongs to the Cannabaceae family, which also includes the genus *Humulus*, known for hops used in brewing (Kovalchuk et al., 2020). Cannabis has been cultivated for thousands of years, with its uses ranging from fiber production to its psychoactive and therapeutic properties (Hurgobin et al., 2020; Romero et al., 2020). The plant is characterized by its production of cannabinoids, terpenes, and other specialized metabolites, which contribute to its diverse applications (Romero et al., 2020).

Cannabis holds significant importance across various sectors. Medically, it is renowned for its therapeutic properties, particularly cannabinoids like cannabidiol (CBD) and tetrahydrocannabinol (THC), which have been documented for their effects on human health (Romero et al., 2020; Singh et al., 2020). The relaxation of legal restrictions in many regions has spurred research into its medicinal potential, leading to advancements in understanding its molecular and genetic pathways (Hurgobin et al., 2020; Adams et al., 2021).

In agriculture, cannabis is valued for its adaptability and the production of hemp, a variety of cannabis grown for its strong fibers used in textiles, bioplastics, and construction materials (Vergara et al., 2016; Romero et al., 2020). Industrially, cannabis is utilized for its seeds, which are a source of nutritious oil and protein, and for bioremediation purposes due to its ability to absorb heavy metals from the soil (Vergara et al., 2016; Adams et al., 2021).

Despite the progress in genome sequencing, current assemblies of the cannabis genome remain incomplete, with significant portions unmapped or poorly annotated. This research aims to address these gaps by leveraging modern genomics technologies to provide a more comprehensive and high-resolution view of the cannabis genome. By integrating multi-omics approaches, including genomics, transcriptomics, and metabolomics, this study seeks to elucidate the complex genetic and biochemical pathways involved in cannabinoid and terpene biosynthesis. The ultimate goal is to facilitate the development of novel cannabis cultivars with optimized traits for medicinal, agricultural, and industrial applications, thereby enhancing the utility and economic value of this multifaceted plant.

2 Historical Background of Cannabis Genomic Studies

2.1 Early genetic studies of Cannabis

The early genetic studies of *Cannabis sativa* L. primarily focused on its cultivation and use for various purposes, including fiber, oil, food, and medicinal properties. Cannabis has been cultivated throughout human history, and selective breeding has produced plants for specific uses, such as high-potency marijuana strains and hemp cultivars for fiber and seed production (Bakel et al., 2011). However, scientific research on cannabis was significantly restricted due to its classification as a narcotic under the Single Convention on Narcotic Drugs of 1961, which limited its production and supply except under license (Hurgobin et al., 2020). Despite these restrictions, early genetic studies laid the groundwork for understanding the basic biology and molecular mechanisms controlling key traits in cannabis.

2.2 Developments leading to genomic sequencing

The relaxation of legislation governing cannabis cultivation for research, medicinal, and recreational purposes in certain jurisdictions has accelerated the development of modern genomics technologies applied to cannabis. This shift has enabled more comprehensive examinations of the cannabis genome, including the use of whole genome sequencing (WGS) and other omics-based methods (Hesami et al., 2020; Hurgobin et al., 2020). The first draft genome sequence of *Cannabis sativa* was reported using short-read sequencing approaches, providing a haploid genome sequence of 534 Mb and a transcriptome of 30 000 genes (Bakel et al., 2011). This development marked a significant milestone in cannabis genomic research, allowing for the systematic analysis of genes involved in cannabinoid biosynthesis and other traits of interest.

2.3 Key milestones in Cannabis genome research

Several key milestones have been achieved in cannabis genome research, including the identification of specific genes and genetic variants associated with cannabinoid biosynthesis. For instance, the exclusive occurrence of Δ 9-tetrahydrocannabinolic acid synthase in marijuana strains and its replacement by cannabidiolic acid synthase in hemp cultivars explains the production of psychoactive THC in marijuana but not in hemp (Bakel et al., 2011). Additionally, recent advances in genome-wide sequencing techniques have enabled the identification of low-frequency genetic variants involved in cannabis dependence, highlighting the potential utility of WGS for understanding the genetic basis of cannabis use disorders (Gizer et al., 2018).

Moreover, the application of multi-omics approaches has provided deeper insights into the molecular mechanisms underlying cannabis traits. These approaches have facilitated the identification of correlations between biological processes and metabolic pathways, aiding in the development of therapeutic marijuana strains with tailored cannabinoid profiles and improved agronomic characteristics (Sirangelo et al., 2022). The integration of genomics, transcriptomics, and metabolomics has thus become a powerful tool for advancing cannabis research and breeding programs.

3 Advancements in Cannabis Genome Sequencing

3.1 Major Cannabis genome sequencing projects

Several significant projects have been undertaken to sequence the genome of Cannabis species. One notable project involved the sequencing of wild-type varieties of *Cannabis sativa* using PacBio single-molecule sequencing and Hi-C technology, resulting in a comprehensive de novo genome assembly (Gao et al., 2020).

Another important effort was a meta-analysis of pooled published genomics data, which highlighted the incomplete nature of current Cannabis genome assemblies and emphasized the need for coordinated efforts to improve the quality and completeness of these assemblies (Kovalchuk et al., 2020). Following these advancements, a groundbreaking study successfully established an *Agrobacterium*-mediated genetic transformation and CRISPR/Cas9-mediated targeted mutagenesis in *Cannabis sativa*, providing a valuable tool for functional genomic studies (Zhang et al., 2021) (Figure 1).

3.2 Methodologies used in sequencing

The methodologies employed in Cannabis genome sequencing have evolved significantly over time. Initially, second-generation sequencing (SGS) technologies were used, but these were limited by short read lengths and difficulties in resolving repetitive regions (Lu et al., 2016). The advent of third-generation sequencing (TGS) technologies, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), has revolutionized the field. PacBio's HiFi reads offer high per-base accuracy and long read lengths, making them suitable for high-quality de novo assemblies (Murigneux et al., 2020; Nurk et al., 2020). ONT's MinION and PromethION platforms provide ultra-long reads, which are beneficial for assembling complex genomes and detecting structural variations (Lu et al., 2016; Jain et al., 2017; Murigneux et al., 2020).

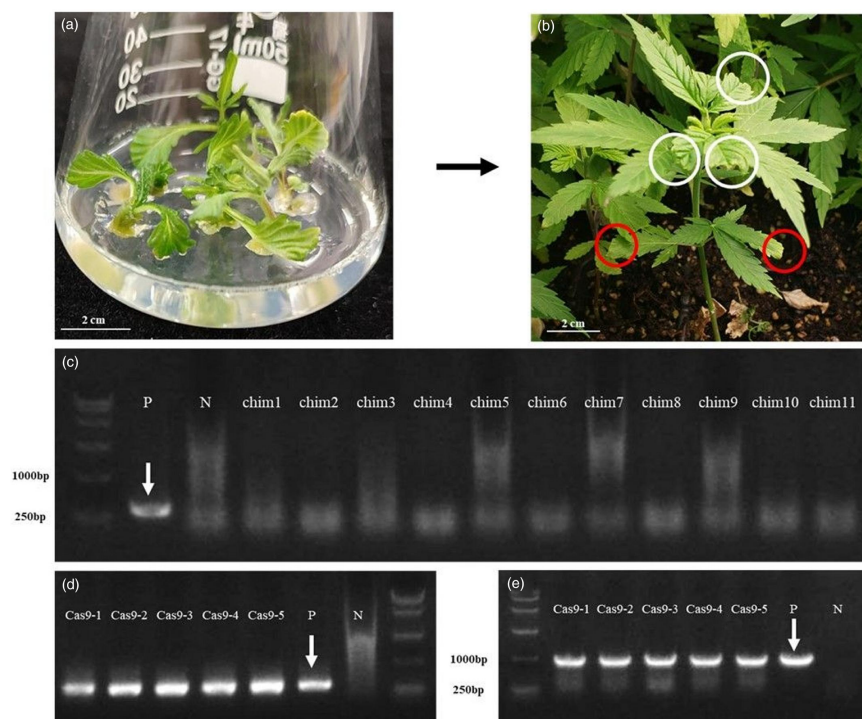


Figure 1 Transgenic cannabis seedlings and results of transgenic screening (Adopted from Zhang et al., 2021)

Image caption: (a) Shoots regenerated from stems of the transgenic seedling. In the estimation of developmental regulator effects on shoot organogenesis, we obtained one transgenic seedling G41-1 carrying the pG41sg T-DNA fragment. Then G41-1 stem was cut into pieces and incubated in the regeneration medium containing kanamycin for 6 weeks. There are five shoots germinated from the stem explant. (b) A transgenic seedling regenerated from the G41-1 stem. All the five shoots were transferred to soil after a 5-week incubation in the root-induction medium. The first fully expanded leaves were sampled every three weeks when growing in greenhouse. Red circle: sampling in the first round of screening; white circle: sampling in the second round of screening. (c) Transgenic-specific PCR result of the chimeric plants containing mutagenesis at CsPDS1. Eleven chimeric seedlings (chim1-11) were randomly selected and transferred to soil after incubation in the root-induction medium. Since their first fully expanded leaves lost the T-DNA fragment, these plants were identified as no transgenic with primers AtU6-F1/R1 in the second round of screening. P: DNA sample of pG41sg, N: DNA sample of no transgenic plant; white arrows: specific PCR product. (d) and (e) Transgenic-specific PCR results of the five seedlings regenerated from G41-1. In the second round of screening, these plants (Cas9-1 to Cas 9-5) were identified as transgenic plants based on transgenic-specific PCR results amplified with primers AtU6-F1/R1 and CsCAS9F2/R2 (Adopted from Zhang et al., 2021)

3.3 Assembly and annotation of Cannabis genomes

The assembly and annotation of Cannabis genomes have seen significant improvements with the use of TGS technologies. For instance, the assembly of wild-type *C. sativa* varieties achieved scaffold and contig N50 sizes of 83.00 Mb and 513.57 kb, respectively, with 98.20% of the protein-coding genes functionally annotated (Gao et al., 2020). The use of hybrid sequencing strategies, combining long reads from TGS with short reads from SGS, has further enhanced the accuracy and completeness of genome assemblies (Rhoads and Au, 2015). Tools like HiCanu have been developed to leverage the high accuracy of PacBio HiFi reads, resulting in superior assembly continuity and accuracy (Nurk et al., 2020). Higher assembly quality is achieved by integrating PacBio SMRT long read length and HiCanu plotting (Wei et al., 2024).

3.4 Comparative genomics of Cannabis species

Comparative genomics of different Cannabis species, such as *Cannabis sativa* and *Cannabis indica*, has provided insights into their evolutionary history and genetic diversity. Studies have shown that current genome assemblies are incomplete, with significant portions of the genome missing or unmapped, which complicates accurate annotation and comparison (Kovalchuk et al., 2020). The use of advanced sequencing technologies and improved assembly methods is essential for generating high-quality reference genomes that can facilitate comparative studies and the identification of species-specific genetic traits (Li et al., 2017; Lang et al., 2020).

4 Functional Gene Mining in Cannabis

4.1 Key genes identified for Cannabinoid biosynthesis

Cannabinoid biosynthesis in *Cannabis sativa* is a complex process involving several key genes. The primary enzymes responsible for the production of major cannabinoids include tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS). These enzymes are crucial for the synthesis of Δ 9-tetrahydrocannabinol (THC) and cannabidiol (CBD), respectively. Studies have shown that the expression levels of these genes vary significantly between different cannabis strains, with THCAS being predominantly expressed in drug-type strains like Purple Kush, while CBDAS is more common in hemp varieties such as 'Finola' (Bakel et al., 2011; Fulvio et al., 2021) (Figure 2). Additionally, the presence of cannabichromenic acid synthase (CBCAS) has been noted, although its role in cannabinoid biosynthesis is less clear and requires further investigation (Fulvio et al., 2021).

4.2 Genes involved in resistance to diseases and environmental stress

Cannabis sativa has evolved various genetic mechanisms to resist diseases and environmental stress. Recent genomic studies have identified several candidate genes associated with these traits. For instance, genes involved in the synthesis of cellulose and lignin have been linked to structural integrity and resistance to pathogens (Ren et al., 2021). Genes related to regulation of salt stress are involved in plant response to salt stress through expression (Liu et al., 2022). Moreover, the identification of single nucleotide polymorphisms (SNPs) in genes related to stress responses provides insights into the plant's ability to adapt to different environmental conditions (Zhao et al., 2021). These genetic markers are crucial for breeding programs aimed at developing disease-resistant and environmentally resilient cannabis strains.

4.3 Functional genes related to fiber production and plant growth

The production of high-quality fiber in hemp varieties of *Cannabis sativa* is governed by specific genes that influence fiber content and plant growth. Research utilizing specific length amplified fragment sequencing (SLAF-seq) and bulked segregant analysis (BSA) has identified several genes that are highly correlated with fiber content. These include genes involved in transcription regulation, auxin transport, and sugar metabolism (Zhao et al., 2021). Additionally, the genetic differentiation between drug-type and fiber-type cannabis has been linked to variations in the THCAS and CBDAS genes, which also affect plant growth and fiber quality (Cascini et al., 2019). The first cannabis spike-type gene CsMIKC1, cloned from cannabis, reveals the molecular mechanism driving the development of female cannabis flowers and is the starting point for elucidating the functions of many homologous genes involved in inflorescence development (Xu et al., 2024).

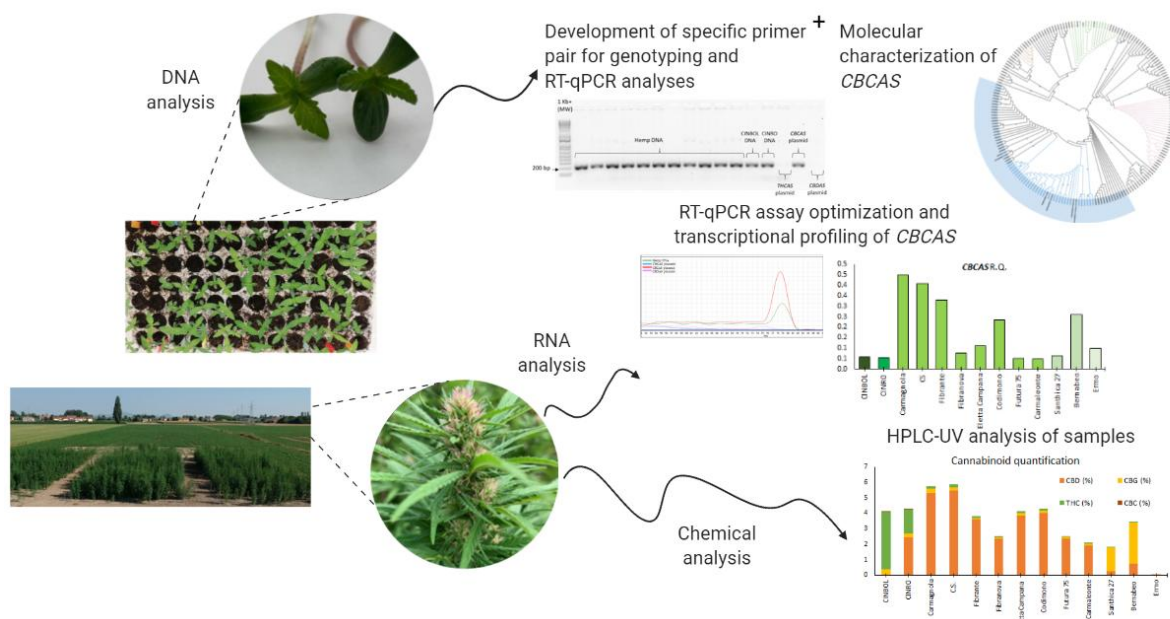


Figure 2 Molecular and chemical characterization of cannabinoid synthase genes in European Cannabis genotypes (Adapted from Fulvio et al., 2021)

4.4 Technological tools for gene mining

Advancements in genomic technologies have significantly enhanced the ability to mine functional genes in *Cannabis sativa*. Techniques such as transcriptomics, proteomics, and next-generation sequencing have been instrumental in identifying and characterizing genes involved in cannabinoid biosynthesis, disease resistance, and fiber production. For example, the use of RNA sequencing (RNA-seq) has enabled the detailed analysis of gene expression profiles in different cannabis strains, revealing key differences in metabolic pathways (Bakel et al., 2011; Romero et al., 2020). Proteomic approaches have further elucidated the diversity of enzymes involved in cannabinoid synthesis and their regulatory mechanisms (Romero et al., 2020). Additionally, the integration of multi-omics data, including genomics, transcriptomics, and metabolomics, has provided a comprehensive understanding of the complex gene networks in cannabis (Wu et al., 2021).

5 Role of Genomic Tools in Breeding Programs

5.1 Use of genomic selection for trait improvement

Genomic selection has become a cornerstone in modern breeding programs, leveraging high-throughput DNA marker genotyping and whole genome sequencing to enhance the selection process. This approach allows breeders to predict the genetic value of plants more accurately and efficiently, thereby accelerating the development of new varieties with desirable traits. The integration of genomic selection with traditional breeding methods has shown significant promise in improving yield, disease resistance, and stress tolerance in various crops (Thomson et al., 2022).

5.2 CRISPR-Cas9 and other gene-editing tools in Cannabis breeding

The advent of CRISPR-Cas9 and other gene-editing technologies has revolutionized plant breeding, including Cannabis. CRISPR-Cas9, in particular, offers a precise, efficient, and relatively simple method for targeted genome modifications. This technology enables the deletion of detrimental traits and the addition of beneficial ones, making it a powerful tool for functional genomics and crop improvement (Bortesi and Fischer, 2015; Arora and Narula, 2017; Jaganathan et al., 2018). The versatility of CRISPR-Cas9 extends to generating knockouts, precise modifications, and multiplex genome engineering, which are crucial for developing Cannabis varieties with enhanced traits (Arora and Narula, 2017; Ahmar et al., 2020; Zhang et al., 2021). Additionally, the development of CRISPR ribonucleoproteins (RNPs) has addressed some limitations of plasmid-based systems, further enhancing the efficiency and applicability of this technology in Cannabis breeding (Arora and Narula, 2017).

5.3 Applications in improving yield, cannabinoid profiles, and environmental adaptability

The application of genomic tools, including CRISPR-Cas9, has significant implications for improving yield, cannabinoid profiles, and environmental adaptability in Cannabis. By enabling precise modifications at the genetic level, these tools can enhance the expression of genes associated with higher yield and better cannabinoid profiles, such as THC and CBD content (Ahmar et al., 2020; Rao and Wang, 2021). Moreover, CRISPR-Cas9 has been used to develop Cannabis varieties with improved resistance to biotic and abiotic stresses, such as pests, diseases, and environmental extremes, thereby increasing the plant's adaptability and overall productivity (Zhang et al., 2017; Jaganathan et al., 2018; Nascimento et al., 2023). The ability to fine-tune gene regulation and create high-throughput mutant libraries further supports the development of Cannabis strains that meet specific agricultural and medicinal needs (Chen et al., 2019; Thomson et al., 2022).

6 Epigenetic Studies in Cannabis

6.1 Understanding epigenetic regulation in Cannabis gene expression

Epigenetic regulation plays a crucial role in the gene expression of Cannabis, influencing various biological processes without altering the DNA sequence. Epigenetic mechanisms such as DNA methylation, histone modifications, and RNA-associated alterations are pivotal in modulating gene expression. For instance, the endocannabinoid system (ECS), which includes cannabinoid receptors and their endogenous ligands, is subject to epigenetic regulation. This regulation can affect the expression of genes involved in neurotransmitter signaling and other critical functions (Basavarajappa and Subbanna, 2022; Bunsick et al., 2023). Additionally, the spatial organization of the cell nucleus and the three-dimensional chromatin architecture are essential for the precise control of gene expression, which can be epigenetically coordinated (Reece and Hulse, 2023).

6.2 Role of epigenetics in phenotype expression, including cannabinoid production

Epigenetic modifications significantly impact phenotype expression in Cannabis, including the production of cannabinoids. These modifications can lead to long-term changes in gene expression that influence the plant's metabolic pathways. For example, DNA methylation and histone modifications can alter the expression of genes involved in cannabinoid biosynthesis, affecting the levels of compounds such as THC and CBD (Wu et al., 2021; Bunsick et al., 2023). Moreover, epigenetic changes can be heritable, potentially influencing the phenotype across generations. This transgenerational inheritance can result from environmental factors, such as exposure to cannabinoids, which can induce epigenetic reprogramming and affect the plant's metabolic phenotype (Bunsick et al., 2023).

6.3 Emerging research in Cannabis epigenomics

Recent advancements in next-generation sequencing technologies have enabled more detailed studies of the Cannabis epigenome. These studies have revealed complex networks of gene regulation involving alternative splicing, microRNAs (miRNAs), and long non-coding RNAs (lncRNAs). For instance, comprehensive transcriptome analyses have identified numerous transcripts encoding key enzymes in cannabinoid biosynthesis, with many of these transcripts undergoing alternative splicing. Additionally, miRNAs and lncRNAs have been shown to target transcripts involved in cannabinoid production, further highlighting the intricate regulatory mechanisms at play (Wu et al., 2021). Emerging single-cell epigenomic methods also hold promise for transforming our understanding of gene regulation and cell identity in Cannabis, offering insights into how epigenetic information is integrated with genomic and transcriptional data (Clark et al., 2016).

7 Case Study: *Cannabis sativa* Genome

7.1 In-depth analysis of *Cannabis sativa* genomic structure

The genomic structure of *Cannabis sativa* has been extensively studied, revealing significant insights into its complexity and diversity. Initial genome assemblies were found to be incomplete, with approximately 10% of the genome missing and 10%-25% unmapped, including critical regions such as ribosomal DNA clusters and centromeres (Kovalchuk et al., 2020). Recent advancements have led to the development of a high-quality reference genome using PacBio single-molecule sequencing and Hi-C technology, resulting in an assembled genome of approximately 808 Mb with a high level of heterozygosity (Gao et al., 2020). This comprehensive genome provides a valuable resource for further genetic and molecular studies.

7.2 Key functional genes and their roles in cannabinoid production

Cannabinoid production in *Cannabis sativa* is primarily governed by specific genes involved in the biosynthesis pathways. The draft genome and transcriptome analysis of the marijuana strain *Purple Kush* identified key genes such as Δ 9-tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS), which are responsible for the production of THC and CBD, respectively (Bakel et al., 2011). These genes are located within large retrotransposon-rich regions that exhibit significant structural differences between drug-type and hemp-type alleles (Laverty et al., 2018). Additionally, the gene encoding cannabichromenic acid synthase (CBCAS) has been characterized, providing further insights into the diversity of cannabinoids produced by different cannabis strains.

7.3 Insights gained from genome sequencing in agricultural and medical applications

Genome sequencing of *Cannabis sativa* has provided critical insights that have significant implications for both agricultural and medical applications. The availability of a high-quality reference genome facilitates the development of cannabis cultivars with tailored cannabinoid profiles, enhancing their therapeutic potential (Bakel et al., 2011). Furthermore, the identification of genetic markers linked to sex determination and cannabinoid content aids in the breeding of cannabis with desired traits, such as increased fiber production or specific medicinal properties (Pan et al., 2021; Ren et al., 2021) (Figure 3). These advancements in genomics are paving the way for more efficient and targeted breeding programs, ultimately contributing to the optimization of cannabis for various uses (Hurgobin et al., 2020; Sirangelo et al., 2022).

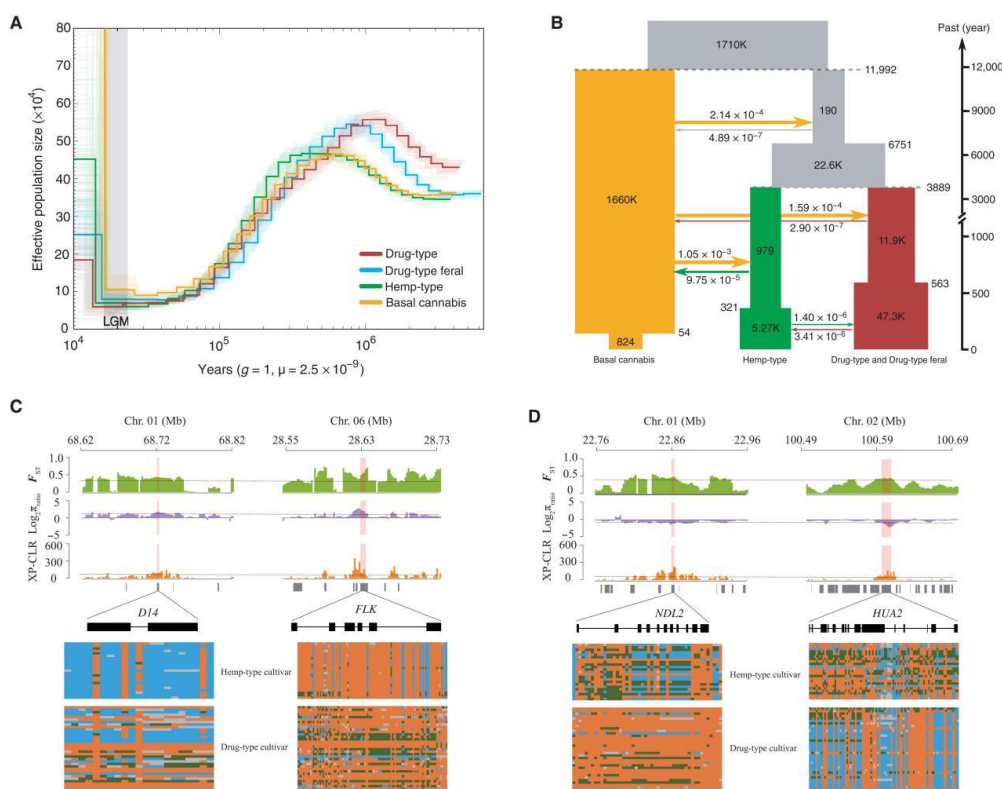


Figure 3 Demographic history of *C. sativa* and selection signatures identified from comparison between hemp- and drug-type cultivars (Adopted from Ren et al., 2021)

Image caption: (A) Demographic history inferred from the PSMC method. (B) Graphical summary of the best-fitting demographic model inferred by fastsimcoal2. Widths show the relative effective population sizes (N_e). Arrows and figures at the arrows indicate the average number of migrants per generation among different groups. The point estimates and 95% confidence intervals of demographic parameters are shown in table S3. Examples of genes with selection sweep signals in hemp-type cultivars (C) and drug-type cultivars (D). Three independent sets of signals (F_{ST} , π ratio, and XP-CLR) are shown along the genomic regions covering the four genes. Dashed lines represent the top 5% of the corresponding values. Below the three plot schemes are the gene models in the genomic regions. Below each gene model are the SNP allele distributions along each of the four genes for the two groups (green, heterozygous site; orange, homozygous site of reference allele; blue, homozygous site of alternative allele; gray, missing data) (Adopted from Ren et al., 2021)

8 Challenges and Limitations

8.1 Limitations in genome sequencing data and accessibility

Despite significant advancements in genome sequencing technologies, the current genome assemblies of *Cannabis sativa* remain incomplete. Approximately 10% of the genome is missing, and 10%-25% remains unmapped. Critical regions such as 45S and 5S ribosomal DNA clusters, centromeres, and satellite sequences are not represented, which hampers the accurate annotation of gene copies (Kovalchuk et al., 2020). Additionally, the high heterozygosity in wild-type varieties of *C. sativa* poses challenges in generating a complete and accurate genome sequence (Gao et al., 2020). The rapid generation of large amounts of sequencing data also raises issues related to data storage, processing, and accessibility, which are critical for advancing genomic research (Kircher and Kelso, 2010; Kahn, 2011).

8.2 Genetic diversity and its impact on functional gene mining

The genetic diversity within *Cannabis sativa*, particularly between its subspecies and cultivars, complicates functional gene mining. The high heterozygosity observed in wild-type varieties introduces a wide range of genetic variations that need to be accounted for in genomic studies (Gao et al., 2020). Single nucleotide polymorphisms (SNPs) in cannabinoid synthase genes significantly affect the plant's chemotype, making it essential to understand these variations for breeding programs aimed at specific cannabinoid profiles (Singh et al., 2020). Moreover, the uncertain taxonomic classification of Cannabis subspecies further complicates the genetic analysis and breeding efforts (Hurgobin et al., 2020).

8.3 Ethical and regulatory issues in Cannabis research

Cannabis research is heavily regulated due to its classification as a narcotic drug under international treaties such as the Single Convention on Narcotic Drugs of 1961. These regulations have historically restricted scientific research and cultivation of cannabis, limiting the availability of genetic resources and hindering progress in genomics studies (Hurgobin et al., 2020). Although some jurisdictions have relaxed these regulations, ethical concerns regarding the use of cannabis for recreational and medicinal purposes continue to pose challenges. Ensuring compliance with varying legal frameworks and addressing societal concerns are critical for the advancement of cannabis research (Hurgobin et al., 2020; Sirangelo et al., 2022).

8.4 Technical challenges in Cannabis cultivation for genomic studies

Cultivating *Cannabis sativa* for genomic studies presents several technical challenges. The dioecious nature of the plant, where male and female flowers develop on separate plants, complicates breeding programs and the study of flower development (Hurgobin et al., 2020). Additionally, the need for controlled growing conditions to ensure consistent phenotypic expression adds to the complexity of cultivation. The high variability in cannabinoid content among different cultivars necessitates precise control over environmental factors to obtain reliable data for genomic studies (Sirangelo et al., 2022). Furthermore, the lack of robust SNP markers and the need for a comprehensive set of SSR markers for marker-assisted breeding programs highlight the technical limitations in current cannabis genomics research (Hurgobin et al., 2020).

9 Future Directions in Cannabis Genomics

9.1 Potential breakthroughs in gene mining and functional genomics

The future of cannabis genomics holds significant promise for breakthroughs in gene mining and functional genomics. One of the key areas of focus is the identification and functional characterization of genes involved in the biosynthesis of cannabinoids and terpenes, which are critical for the plant's medicinal properties. Recent studies have demonstrated the potential of virus-induced gene silencing (VIGS) to knock down specific genes in *Cannabis sativa*, providing a powerful tool for reverse genetic studies to uncover unknown gene functions (Schachtsiek et al., 2019). Additionally, advancements in in silico analysis and genome editing technologies, such as CRISPR, Zinc Fingers, and TALENs, are paving the way for precise modifications of genes involved in cannabinoid biosynthesis, which could lead to the development of new cannabis strains with enhanced therapeutic properties (Matchett-Oates et al., 2021).

9.2 Future technologies for sequencing and gene analysis

The rapid evolution of sequencing technologies is expected to further revolutionize cannabis genomics. High-throughput sequencing methods, such as PacBio single-molecule sequencing and Hi-C technology, have already been employed to generate high-quality reference genomes for wild-type varieties of *Cannabis sativa*, providing a comprehensive genetic resource for future research (Gao et al., 2020). These technologies enable the assembly of more complete and accurate genomes, which are essential for detailed genetic and functional analyses. Moreover, the integration of whole genome sequencing (WGS) with advanced analytical methods allows for the identification of low-frequency genetic variants associated with traits such as cannabis dependence, offering new insights into the genetic basis of complex traits (Gizer et al., 2018).

9.3 Integration of multi-omics approaches for Cannabis research

The integration of multi-omics approaches, including genomics, transcriptomics, proteomics, and metabolomics, is poised to significantly enhance our understanding of cannabis biology and its applications. Omics-based methods have already been utilized to study the molecular markers, microRNAs, and functional genes related to terpene and cannabinoid biosynthesis, as well as fiber quality in *Cannabis sativa* (Hesami et al., 2020). By combining data from multiple omics layers, researchers can gain a holistic view of the regulatory networks and metabolic pathways involved in the production of bioactive compounds. This comprehensive approach will facilitate the identification of key regulatory genes and pathways, ultimately leading to the development of improved cannabis cultivars with optimized traits for medicinal, industrial, and agricultural use (Vergara et al., 2016; Hurgobin et al., 2020; Adams et al., 2021).

10 Concluding Remarks

Recent advancements in the field of Cannabis genomics have significantly enhanced our understanding of this multifaceted plant. The sequencing of the *Cannabis sativa* genome has revealed a complex genetic structure with substantial heterozygosity and a high level of genetic variation among different cultivars. Despite these advancements, current genome assemblies remain incomplete, with notable gaps and low-resolution ordering, which complicates the accurate annotation of genes. The application of multi-omics approaches, including genomics, transcriptomics, and metabolomics, has provided deeper insights into the molecular mechanisms underlying cannabinoid biosynthesis and other traits of interest. Additionally, *in silico* analyses have facilitated the design of genome editing tools, although technical challenges persist due to the highly polymorphic nature of the Cannabis genome.

The progress in Cannabis genomics holds significant implications across various fields. In medicine, the ability to tailor cannabinoid profiles through genomic insights can lead to the development of therapeutic strains with specific medicinal properties. Industrial applications benefit from the genetic improvement of hemp cultivars for fiber and seed production, enhancing their agronomic traits. In agriculture, understanding the genetic diversity and biochemical pathways of Cannabis can aid in breeding programs aimed at improving yield, disease resistance, and environmental adaptability. It also provides a theoretical basis for the synthetic biology of rare cannabinoids. The integration of biotechnological techniques, such as virus-induced gene silencing (VIGS) and genetic engineering, further expands the potential for functional gene studies and the production of high-value metabolites.

To fully unlock the potential of Cannabis genomics, several key areas require further exploration. First, achieving high-quality, complete genome assemblies is essential to close existing gaps and enhance the resolution of genomic data. In addition, the use of functional genomics tools, such as CRISPR and virus-induced gene silencing (VIGS), should be expanded to clarify the gene functions and regulatory networks involved in cannabinoid biosynthesis and other metabolic pathways. The integration of multi-omics approaches is also critical, as it enables the correlation of genotypic and phenotypic data, offering a comprehensive understanding of the molecular mechanisms driving important traits. Furthermore, conducting post-culture analyses of Cannabis phytochemistry and pharmacology will ensure the integrity and efficacy of *in vitro* propagated plants, particularly for pharmaceutical applications. Lastly, developing advanced breeding programs that leverage genomic data will

allow for the creation of cultivars with optimized traits tailored to industrial, medicinal, and agricultural needs. By addressing these research areas, the scientific community can drive significant advancements in Cannabis genomics, opening new avenues for innovation and applications across diverse fields.

Acknowledgments

We sincerely thank Professor Qi Xingjiang of Zhejiang Academy of Agricultural Sciences for his help and support in this project, we would also like to extend my sincere thanks to two anonymous peer reviewers for their thorough assessment and constructive comments, which have all contributed significantly to the improvement of this manuscript.

Funding

This paper was funded by the project "Construction of precision Breeding Facilities for Industrial Hemp" (10402110120AP2201F) funded by the special financial fund of Zhejiang Academy of Agricultural Sciences.

Conflict of Interest Disclosure

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

References

- Adams T., Masondo N., Malatsi P., and Makunga N., 2021, *Cannabis sativa*: from therapeutic uses to micropropagation and beyond, *Plants*, 10(10): 2078.
<https://doi.org/10.3390/plants10102078>
- Ahmar S., Saeed S., Khan M., Khan S., Mora-Poblete F., Kamran M., Faheem A., Maqsood A., Rauf M., Saleem S., Hong W., and Jung K., 2020, A revolution toward gene-editing technology and its application to crop improvement, *International Journal of Molecular Sciences*, 21(16): 5665.
<https://doi.org/10.3390/ijms21165665>
- Arora L., and Narula A., 2017, Gene editing and crop improvement using CRISPR-Cas9 system, *Frontiers in Plant Science*, 8: 1932.
<https://doi.org/10.3389/fpls.2017.01932>
- Bakel H., Stout J., Coté A., Tallon C., Sharpe A., Hughes T., and Page J., 2011, The draft genome and transcriptome of *Cannabis sativa*, *Genome Biology*, 12: R102-R102.
<https://doi.org/10.1186/gb-2011-12-10-r102>
- Basavarajappa B., and Subbanna S., 2022, Molecular insights into epigenetics and cannabinoid receptors, *Biomolecules*, 12(11): 1560.
<https://doi.org/10.3390/biom12111560>
- Bortesi L., and Fischer R., 2015, The CRISPR/Cas9 system for plant genome editing and beyond, *Biotechnology Advances*, 33(1): 41-52.
<https://doi.org/10.1016/j.biotechadv.2014.12.006>
- Bunsick D., Matsukubo J., and Szewczuk M., 2023, Cannabinoids Transmogrify cancer metabolic phenotype via epigenetic reprogramming and a novel CBD biased G protein-coupled receptor signaling platform, *Cancers*, 15(4): 1030.
<https://doi.org/10.3390/cancers15041030>
- Cascini F., Farcomeni A., Migliorini D., Baldassarri L., Boschi I., Martello S., Amaducci S., Lucini L., and Bernardi J., 2019, Highly predictive genetic markers distinguish drug-type from fiber-type *Cannabis sativa* L, *Plants*, 8(11): 496.
<https://doi.org/10.3390/plants8110496>
- Chen K., Wang Y., Zhang R., Zhang H., and Gao C., 2019, CRISPR/Cas genome editing and precision plant breeding in agriculture, *Annual Review of Plant Biology*, 70: 667-697.
<https://doi.org/10.1146/annurev-arplant-050718-100049>
- Clark S., Lee H., Smallwood S., Kelsey G., and Reik W., 2016, Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity, *Genome Biology*, 17: 1-10.
<https://doi.org/10.1186/s13059-016-0944-x>
- Fulvio F., Paris R., Montanari M., Citti C., Cilento V., Bassolino L., Moschella A., Alberti I., Pecchioni N., Cannazza G., and Mandolino G., 2021, Analysis of sequence variability and transcriptional profile of *Cannabinoid synthase* genes in *Cannabis sativa* L. chemotypes with a focus on *Cannabichromenic acid synthase*, *Plants*, 10(9): 1857.
<https://doi.org/10.3390/plants10091857>
- Gao S., Wang B., Xie S., Xu X., Zhang J., Pei L., Yu Y., Yang W., and Zhang Y., 2020, A high-quality reference genome of wild *Cannabis sativa*, *Horticulture Research*, 7: 73.
<https://doi.org/10.1038/s41438-020-0295-3>
- Gizer I., Bizon C., Gilder D., Ehlers C., and Wilhelmsen K., 2018, Whole genome sequence study of cannabis dependence in two independent cohorts, *Addiction Biology*, 23: 461-473.
<https://doi.org/10.1111/adb.12489>

- Hesami M., Pepe M., Alizadeh M., Rakei A., Baiton A., and Jones A., 2020, Recent advances in cannabis biotechnology, *Industrial Crops and Products*, 158: 113026.
<https://doi.org/10.1016/j.indcrop.2020.113026>
- Hurgobin B., Tamiru-Oli M., Welling M., Doblin M., Bacic A., Whelan J., and Lewsey M., 2020, Recent advances in *Cannabis sativa* genomics research, *The New Phytologist*, 230: 73-89.
<https://doi.org/10.1111/nph.17140>
- Jaganathan D., Ramasamy K., Sellamuthu G., Jayabalan S., and Venkataraman G., 2018, CRISPR for crop improvement: an update review, *Frontiers in Plant Science*, 9: 985.
<https://doi.org/10.3389/fpls.2018.00985>
- Jain M., Koren S., Miga K., Quick J., Rand A., Sasani T., Tyson J., Beggs A., Dilthey A., Fiddes I., Malla S., Marriott H., Nieto T., O'Grady J., Olsen H., Pedersen B., Rhie A., Richardson H., Quinlan A., Snutch T., Tee L., Paten B., Phillippy A., Simpson J., Loman N., and Loose M., 2017, Nanopore sequencing and assembly of a human genome with ultra-long reads, *Nature Biotechnology*, 36: 338-345.
<https://doi.org/10.1038/nbt.4060>
- Kahn S., 2011, On the future of genomic data, *Science*, 331: 728-729.
<https://doi.org/10.1126/science.1197891>
- Kircher M., and Kelso J., 2010, High-throughput DNA sequencing - concepts and limitations, *BioEssays*, 32(6): 524-536.
<https://doi.org/10.1002/bies.200900181>
- Kovalchuk I., Pellino M., Rigault P., Velzen R., Ebersbach J., Ashnest J., Mau M., Schranz M., Alcorn J., Laprairie R., Laprairie R., McKay J., Burbridge C., Schneider D., Vergara D., Kane N., and Sharbel T., 2020, The genomics of cannabis and its close relatives, *Annual Review of Plant Biology*, 71(1): 713-739.
<https://doi.org/10.1146/annurev-arplant-081519-040203>
- Lang D., Zhang S., Ren P., Liang F., Sun Z., Meng G., Tan Y., Li X., Lai Q., Han L., Wang D., Hu F., Wang W., and Liu S., 2020, Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore, *GigaScience*, 9(12): g1aa123.
<https://doi.org/10.1093/gigascience/g1aa123>
- Laverty K., Stout J., Sullivan M., Shah H., Gill N., Holbrook L., Deikus G., Sebra R., Hughes T., Page J., and Bakel H., 2018, A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci, *Genome Research*, 29: 146-156.
<https://doi.org/10.1101/gr.242594.118>
- Li C., Lin F., An D., Wang W., and Huang R., 2017, Genome sequencing and assembly by long reads in plants, *Genes*, 9(1): 6.
<https://doi.org/10.3390/genes9010006>
- Liu H., Hu H., Tang K., Rehman M., Du G., Huang Y., and Liu F., 2022, Overexpressing hemp salt stress induced transcription factor genes enhances tobacco salt tolerance, *Industrial Crops and Products*, 177: 114497.
<https://doi.org/10.1016/j.indcrop.2021.114497>
- Lu H., Giordano F., and Ning Z., 2016, Oxford nanopore MinION sequencing and genome assembly, *Genomics, Proteomics and Bioinformatics*, 14: 265-279.
<https://doi.org/10.1016/j.gpb.2016.05.004>
- Matchett-Oates L., Braich S., Spangenberg G., Rochfort S., and Cogan N., 2021, In silico analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation, *PLoS ONE*, 16(9): e0257413.
<https://doi.org/10.1371/journal.pone.0257413>
- Murigneux V., Rai S., Furtado A., Bruxner T., Tian W., Ye Q., Wei H., Yang B., Harliwong I., Anderson E., Mao Q., Drmanac R., Wang O., Peters B., Xu M., Wu P., Topp B., Coin L., and Henry R., 2020, Comparison of long-read methods for sequencing and assembly of a plant genome, *GigaScience*, 9(12): g1aa146.
<https://doi.org/10.1093/gigascience/g1aa146>
- Nascimento F., Rocha A., Soares J., Mascarenhas M., Ferreira M., Lino L., Ramos A., Diniz L., Mendes T., Ferreira C., Santos-Serejo J., and Amorim E., 2023, Gene editing for plant resistance to abiotic factors: a systematic review, *Plants*, 12(2): 305.
<https://doi.org/10.3390/plants12020305>
- Nurk S., Walenz B., Rhie A., Vollger M., Logsdon G., Grothe R., Miga K., Eichler E., Phillippy A., and Koren S., 2020, HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads, *Genome Research*, 30: 1291-1305.
<https://doi.org/10.1101/gr.263566.120>
- Pan G., Li Z., Huang S., Tao J., Shi Y., Chen A., Li J., Tang H., Chang L., Deng Y., Li D., and Zhao L., 2021, Genome-wide development of insertion-deletion (InDel) markers for Cannabis and its uses in genetic structure analysis of Chinese germplasm and sex-linked marker identification, *BMC Genomics*, 22: 1-12.
<https://doi.org/10.1186/s12864-021-07883-w>
- Reece A., and Hulse G., 2023, Perturbation of 3D nuclear architecture, epigenomic dysregulation and aging, and cannabinoid synaptopathy reconfigures conceptualization of cannabinoid pathophysiology: part 1-aging and epigenomics, *Frontiers in Psychiatry*, 14: 1182535.
<https://doi.org/10.3389/fpsy.2023.1182535>
- Ren G., Zhang X., Li Y., Ridout K., Serrano-Serrano M., Yang Y., Liu A., Ravikanth G., Nawaz M., Mumtaz A., Salamin N., and Fumagalli L., 2021, Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*, *Science Advances*, 7(29): eabg2286.
<https://doi.org/10.1126/sciadv.abg2286>

- Rhoads A., and Au K., 2015, PacBio sequencing and its applications, *Genomics, Proteomics and Bioinformatics*, 13: 278-289.
<https://doi.org/10.1016/j.gpb.2015.08.002>
- Romero P., Peris A., Vergara K., and Matus J., 2020, Comprehending and improving cannabis specialized metabolism in the systems biology era, *Plant Science : An International Journal of Experimental Plant Biology*, 298: 110571.
<https://doi.org/10.1016/j.plantsci.2020.110571>
- Schachtsiek J., Hussain T., Azzouhri K., Kayser O., and Stehle F., 2019, Virus-induced gene silencing (VIGS) in *Cannabis sativa* L., *Plant Methods*, 15: 1-9.
<https://doi.org/10.1186/s13007-019-0542-5>
- Singh A., Bilichak A., and Kovalchuk I., 2020, The genetics of Cannabis - genomic variations of key synthases and their effect on cannabinoids content, *Genome*, 64(4): 490-501.
<https://doi.org/10.1139/gen-2020-0087>
- Sirangelo T., Ludlow R., and Spadafora N., 2022, Multi-omics approaches to study molecular mechanisms in *Cannabis sativa*, *Plants*, 11(16): 2182.
<https://doi.org/10.3390/plants11162182>
- Thomson M., Biswas S., Tsakirpaloglou N., and Septiningsih E., 2022, Functional allele validation by gene editing to leverage the wealth of genetic resources for crop improvement, *International Journal of Molecular Sciences*, 23(12): 6565.
<https://doi.org/10.3390/ijms23126565>
- Vergara D., Baker H., Clancy K., Keepers K., Mendieta J., Pauli C., Tittes S., White K., and Kane N., 2016, Genetic and genomic tools for *Cannabis sativa*, *Critical Reviews in Plant Sciences*, 35: 364-377.
<https://doi.org/10.1080/07352689.2016.1267496>
- Wei H.W., Yang Z.Q., Niyitanga S., Tao A., Xu J.T., Fang P.P., Lin L.H., Zhang L.M., Qi J.M., Ming R., and Zhang L.W., 2024, The reference genome of seed hemp (*Cannabis sativa*) provides newinsights into fatty acid and vitamin E synthesis, *Plant Communications*, 5(1): 100718.
<https://doi.org/10.1016/j.xplc.2023.100718>
- Wu B., Li Y., Li J., Xie Z., Luan M., Gao C., Shi Y., and Chen S., 2021, Genome-wide analysis of alternative splicing and non-coding RNAs reveal complicated transcriptional regulation in *Cannabis sativa* L, *International Journal of Molecular Sciences*, 22(21): 11989.
<https://doi.org/10.3390/ijms222111989>
- Xu G.C., Liu Y.B., Yu S.H., Kong D.J., Tang K.L., Dai Z.G., Sun J., Cheng C.H., Deng C.H., Yang Z.M., Tang Q., Li C., Su J.G., and Zhang X.Y., 2024, CsMIKC1 regulates inflorescence development and grain production in *Cannabis sativa* plants, *Horticulture Research*, 11(8): uhae161.
<https://doi.org/10.1093/hr/uhae161>
- Zhang H., Zhang J., Lang Z., Botella J., and Zhu J., 2017, Genome editing-principles and applications for functional genomics research and crop improvement, *Critical Reviews in Plant Sciences*, 36: 291-309.
<https://doi.org/10.1080/07352689.2017.1402989>
- Zhang X.Y., Xu G.C., Cheng C.H., Lei L., Sun J., Xu Y., Deng C.H., Dai Z.G., Yang Z.M., Chen X.J., Liu C., Tang Q., and Su J.G., 2021, Establishment of an *Agrobacterium*-mediated genetic transformation and CRISPR/Cas9-mediated targeted mutagenesis in Hemp (*Cannabis sativa* L.), *Plant Biotechnology Journal*, 19(10): 1979-1987.
<https://doi.org/10.1111/pbi.13611>
- Zhao Y., Sun Y., Cao K., Zhang X., Bian J., Han C., Jiang Y., Xu L., and Wang X., 2021, Combined use of specific length amplified fragment sequencing (SLAF-seq) and bulked segregant analysis (BSA) for rapid identification of genes influencing fiber content of hemp (*Cannabis sativa* L.), *BMC Plant Biology*, 22(1): 250.
<https://doi.org/10.1186/s12870-022-03594-w>

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.
