Systematic Review Open Access

Unlocking the Tea Genome: Advances in High-Quality Sequencing and Annotation

Xi Chen [1], Yichen Zhao [2]

1 Guizhou Institute of Pratacultural / Plant Conservation & Breeding Technology Center, Guizhou Academy of Agricultural Sciences, Guiyang, 550006, Guizhou, China.

2 Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), College of Tea Sciences, Guizhou University, Guiyang, 550025, Guizhou, China

Corresponding author: yczhao@gzu.edu.cn

Journal of Tea Science Research, 2024, Vol.14, No.2 doi: 10.5376/jtsr.2024.14.0008

Received: 18 Jan., 2024

Accepted: 20 Feb., 2024

Published: 02 Mar., 2024

Preferred citation for this article:

Chen X., and Zhao Y.C., 2024, Unlocking the tea genome: advances in high-quality sequencing and annotation, Journal of Tea Science Research, 14(2): 79-91 (doi: 10.5376/jtsr.2024.14.0008)

Abstract This study explores the latest advancements in high-quality sequencing and annotation of the tea plant genome, revealing its genetic diversity, regulatory mechanisms, and biotechnological applications. Through comprehensive genomic analysis, significant discoveries have been made, including the assembly of the complex tea plant genome, key genes regulating the synthesis of bioactive compounds (such as catechins and caffeine), and epigenetic regulatory mechanisms influencing tea plant phenotypes and environmental adaptability. Comparative genomics studies have elucidated the relationships between tea cultivars and their wild relatives, enhancing the understanding of genetic variation and adaptive traits. These findings highlight the potential of tea genomics in precision breeding, which can be used to develop climate-resistant cultivars, improve tea quality, and diversify market products. The advancements in high-quality sequencing and annotation of the tea plant genome have significantly improved our understanding of the genetic and metabolic bases of tea quality. These discoveries provide valuable resources for future research and breeding programs aimed at improving tea plant varieties and expanding the diversity of tea flavors.

Keywords Tea plant genome; High-quality sequencing; Annotation; Genetic diversity; Biotechnological applications; Precision breeding

1 Introduction

Tea (*Camellia sinensis*) is one of the most widely consumed beverages globally, cherished not only for its unique flavors but also for its numerous health benefits. The economic, cultural, and medicinal significance of tea has driven extensive research into its genetic makeup. Understanding the tea genome is crucial as it provides insights into the genetic basis of key quality traits of tea, such as flavor, aroma, and health-promoting compounds (Shi et al., 2011; Xia et al., 2017; Wei et al., 2018). Simultaneously, it aids in the identification of genes responsible for stress resistance and adaptation, which is vital for improving tea crop resilience in the face of climate change (Xia et al., 2020a; Kong et al., 2022). Furthermore, comprehensive genomic knowledge facilitates advanced breeding programs aimed at developing superior tea varieties with enhanced traits (Zhang et al., 2021a).

The journey of tea genome sequencing has seen significant milestones over the past years. Early efforts were marked by the draft genome sequence of *Camellia sinensis* var. *sinensis*, which provided initial insights into the evolution of the tea genome and its quality traits (Wei et al., 2018). Subsequent studies focused on high-quality genome assemblies, such as the haplotype-resolved genome assembly of the Oolong tea cultivar Tieguanyin, which shed light on the evolutionary history and genetic diversity of tea plants (Zhang et al., 2021a). The reference genome of *Camellia sinensis* var. *sinensis*, consisting of 15 pseudo-chromosomes, further elucidated the roles of LTR retrotransposons in genome size expansion and gene diversification (Xia et al., 2020a). Additionally, transcriptome profiling using advanced sequencing technologies has revealed candidate genes for major metabolic pathways, enhancing our understanding of tea-specific compounds (Shi et al., 2011; Wang et al., 2020).

This study synthesizes existing knowledge and integrates recent advancements in tea genome sequencing and annotation methods to identify emerging trends, challenges, and future directions in the field. It explores the latest methods for generating high-quality reference genomes, strategies for comprehensive annotation through the integration of multi-omics data, and the impact of genomic insights on tea breeding and biotechnological applications. Additionally, it elucidates the evolutionary mechanisms shaping the tea genome, including whole-genome duplications and lineage-specific gene expansions. The study aims to provide valuable genomic resources to facilitate future functional genomic research and breeding programs, ultimately contributing to the development of improved tea varieties with desirable traits. By achieving these goals, the study will significantly advance the understanding of the tea genome and provide a theoretical foundation for innovations in tea cultivation and production.

2 Advances in Sequencing Technologies

2.1 Overview of sequencing technologies

The field of genomics has seen significant advancements in sequencing technologies over the past few decades. Initially, Sanger sequencing provided the foundational framework for early genome projects, including the initial draft sequences of the tea genome. However, the high cost and labor-intensive nature of Sanger sequencing limited its scalability for large-scale genomic studies. The emergence of next-generation sequencing (NGS) technologies marked a turning point in genomic research. NGS methods have replaced traditional Sanger sequencing, with NGS platforms providing unprecedented throughput by generating millions of short-read sequences in parallel at a lower cost per base (Bansal et al., 2018).

Among the most prominent NGS platforms are Illumina and PacBio, both of which have been instrumental in sequencing complex genomes such as that of the tea plant, *Camellia sinensis* (Xia et al., 2017; Wei et al., 2018; Xia et al., 2020). Illumina sequencing, known for its high accuracy and short read lengths, and PacBio sequencing, which provides longer reads albeit at a higher error rate, have been used in tandem to achieve comprehensive genome assemblies (Wei et al., 2018).

2.2 Recent developments in high-throughput sequencing

Recent years have witnessed remarkable advancements in high-throughput sequencing technologies tailored to meet the specific challenges of tea genome research. Recent developments in high-throughput sequencing (HTS) have further enhanced our ability to decode complex genomes. For instance, the use of single sperm sequencing has enabled the phasing of highly heterozygous genomes, providing insights into genetic recombination and allele-specific expression (Zhang et al., 2020a). Additionally, the integration of genotyping-by-sequencing (GBS) methods has facilitated the discovery of single nucleotide polymorphisms (SNPs) across diverse tea cultivars, aiding in the understanding of genetic control over desirable traits (Hazra et al., 2020). Comparative studies of sequencing platforms, such as Illumina and Ion Torrent, have also been conducted to optimize protocols for food quality control and component identification in herbal teas (Speranskaya et al., 2018).

2.3 Advantages of modern sequencing methods

Modern sequencing methods offer several advantages over traditional techniques. High-throughput platforms like Illumina and PacBio allow for the rapid and cost-effective sequencing of large genomes, providing high-resolution genetic maps and facilitating the identification of key genes involved in important traits such as flavor, stress resistance, and metabolite production (Xia et al., 2017; Wei et al., 2018; Xia et al., 2020a). The ability to generate phased haploid genomes from diploid organisms enhances functional genomic studies and breeding programs by revealing complex genetic relationships and crossover patterns (Zhang et al., 2020a). Furthermore, the discovery of SNP markers through GBS methods enables precise trait association studies, which are crucial for crop improvement (Hazra et al., 2020). These advancements collectively contribute to a deeper understanding of the tea genome and pave the way for future research and breeding efforts aimed at enhancing tea quality and diversity.

Continuous innovation in sequencing technologies has led to unprecedented advancements in tea genome research, opening new dimensions in genetic diversity, functional genomics, and biotechnological applications. These advancements collectively enhance the understanding of the tea genome, enabling researchers to utilize genomic resources for precision breeding strategies, develop tea plant varieties resistant to biotic and abiotic stresses, and improve tea quality and nutritional value.

3 Genome Assembly Challenges

3.1 Complexities in tea genome

The assembly of the tea genome faces significant challenges due to its inherent complexity, including large genome size, abundant repetitive sequences, and high heterozygosity. The tea genome size typically ranges from 2.5 to 4.0 gigabases (Gb), depending on the species and ploidy level. This large genome size poses challenges for sequencing, assembly, and subsequent annotation processes. Additionally, the high proportion of repetitive sequences in the tea genome, accounting for at least 64%, further complicates assembly (Wei et al., 2018). These repetitive sequences can lead to misassemblies and gaps in the final genome sequence.

The high degree of heterozygosity and structural variation within the genome adds another layer of complexity to the assembly process (Chin et al., 2016). This introduces allelic variation and haplotype diversity that must be resolved during genome reconstruction. These complexities highlight the need for robust computational algorithms and innovative sequencing strategies to achieve high-quality, contiguous assemblies essential for downstream genomic analysis and biotechnological applications. Furthermore, the tea genome has undergone multiple whole-genome duplications, resulting in indistinguishable homologous gene copies, adding another layer of complexity to the assembly (Wei et al., 2018).

3.2 Tools and techniques for effective assembly

To address these complexities, various tools and techniques have been developed and optimized for effective genome assembly. Bioinformatics pipelines, such as de novo assemblers (e.g., SOAPdenovo, SPAdes) and reference-guided assembly tools (e.g., BWA-MEM, Bowtie2), play pivotal roles in reconstructing tea genome sequences from raw sequencing data. Long-read sequencing technologies, such as those provided by Oxford Nanopore's MinION and PacBio, have become essential for spanning repetitive regions and achieving more contiguous assemblies (Chin et al., 2016; Mgwatyu et al., 2022). Optical mapping and chromosome conformation capture techniques provide complementary structural information, facilitating the scaffolding and validation of assembled genome sequences.

The FALCON and FALCON-Unzip algorithms are particularly effective for assembling highly heterozygous genomes, as they can generate phased diploid assemblies that accurately represent the haplotype structure (Chin et al., 2016). The study found that FALCON-Phase addresses phase-switching issues by using Hi-C short-read sequences, thereby reconstructing long haplotype blocks. This method has demonstrated high accuracy (>96%) in benchmark tests (Kronenberg et al., 2018). Additionally, the Bridger package has shown superior performance in transcriptome assembly, providing high completeness and accuracy, which is crucial for understanding gene expression and regulation in tea plants (Li et al., 2019).

3.3 Case studies of successful genome assemblies

Several successful genome assembly projects have been conducted on tea plants and related species, providing valuable insights and resources for the research community. For instance, Xia et al. (2020a) assembled the genome of *Camellia sinensis* var. *sinensis* using a combination of Single-Molecule Real-Time (SMRT) sequencing and Chromosome Conformation Capture (Hi-C) technology, resulting in a high-quality genome assembly (Figure 1). Through comparative genomics, phylogenetics, transcriptomics, and population genetics analyses, they gained deep insights into the evolution and adaptability of the tea tree genome. This study produced a highly contiguous assembly, improving the resolution of repetitive elements and gene-rich regions, enabling comprehensive gene annotation and comparative genomic analysis between tea varieties.

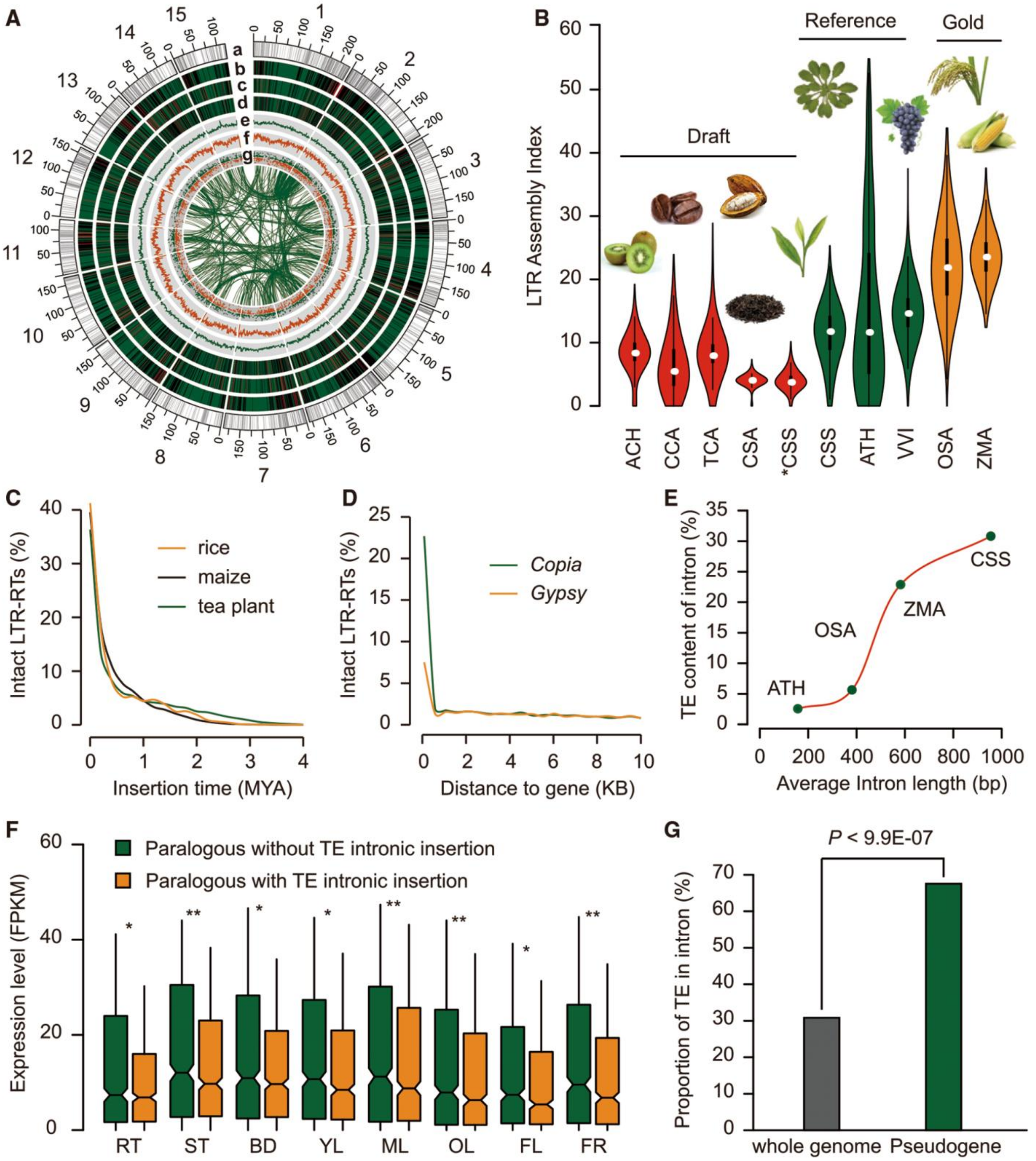

Figure 1 Genome Assembly and Evolution of LTR-RTs of the Tea Plant (Cultivar Shuchazao) (Adopted from Xia et al., 2020a)

Image caption: A: The multi-layer landscape of the tea plant genome, including transcription factor density, gene density, distribution of LTR-RTs, and SSR density; B: Comparison of LAI assessments between two tea plant varieties and other plants, indicating the high assembly quality of the tea genome; C: Insertion time of LTR-RTs in tea plants, rice, and maize, revealing that most LTR-RTs in the tea genome were inserted within the last 1 million years; D: Distribution of distances between LTR-RTs and protein-coding genes in tea plants, showing that most LTR-RTs are located within 2 kb upstream of genes, potentially playing a regulatory role in gene expression (Adapted from Xia et al., 2020a)

Wei et al. (2018) used Illumina and PacBio sequencing technologies to obtain high-quality tea genome sequences. The results showed that the tea genome underwent two whole-genome duplications (WGD), occurring approximately 30-40 million years ago and 90-100 million years ago. Genome duplication and subsequent gene amplification significantly affected the copy number of secondary metabolite genes, particularly those crucial for tea quality, including catechins, theanine, and caffeine synthesis genes. Through transcriptome and phytochemical data analysis, key gene families related to unique tea metabolites were found to have undergone expansion and transcriptional differentiation. This genome sequence aids in the in-depth understanding of tea genome evolution and metabolic pathways, promoting the improvement of tea breeding.

Another study successfully assembled the genome of a wild tea tree, DASZ, at a chromosome scale, which helped clarify the pedigree and selection history of tea varieties and identified key genes involved in flavonoid biosynthesis (Zhang et al., 2020). Additionally, the genome of rooibos (*Aspalathus linearis*), an important medicinal plant used for tea production, was assembled using various long-read sequencing approaches, demonstrating the effectiveness of these techniques in generating contiguous and accurate genome assemblies (Mgwatyu et al., 2022).

4 Annotation and Functional Analysis

4.1 Importance of accurate annotation

Accurate annotation of the tea genome is crucial for understanding the genetic basis of important traits and for facilitating breeding programs. Annotation helps in identifying gene functions, regulatory elements, and structural variations, which are essential for functional genomics studies. For instance, the identification and characterization of unigene derived microsatellite (UGMS) markers in tea have provided insights into the genetic diversity and heterozygosity of tea populations, which are vital for genetic mapping and marker-assisted selection (Ou et al., 2019). Moreover, accurate annotation aids in the identification of gene family members critical for the biosynthesis of key tea metabolites, such as catechins and theanine, which contribute to tea quality and its health benefits (Wei et al., 2018).

4.2 Methods of gene annotation

Gene annotation in tea involves several methods, including the use of high-throughput sequencing technologies and bioinformatics tools. The draft genome sequence of *Camellia sinensis* var. *sinensis* was assembled using both Illumina and PacBio sequencing technologies, which provided a high-quality genome assembly with 33,932 high-confidence predictions of encoded proteins (Wei et al., 2018). Additionally, functional annotation of unigenes containing SSRs was performed through gene ontology (GO) characterization, revealing significant sequence similarity with known proteins in *Arabidopsis thaliana* (Ou et al., 2019). Single sperm sequencing has also been employed to phase the genome of tea, aiding in the construction of high-resolution genetic and recombination maps (Zhang et al., 2020a).

4.3 Insights from functional genomics

Functional genomics studies have provided valuable insights into the regulation of gene expression and the evolution of the tea genome. For example, By sequencing the genomes of 135 single sperm cells, researchers successfully phased the genome of the 'Fudingdabai' tea plant and constructed a high-resolution genetic map, revealing the distribution patterns and interference mechanisms of crossover sites (Figure 2). The study found that crossover locations were often at the 5' and 3' ends of genes and were distributed relatively randomly. Additionally, the researchers developed a method to infer kinship among tea germplasm, detecting complex kinship and genetic markers, which are crucial for understanding genetic diversity and breeding strategies (Zhang et al., 2020a). Furthermore, the draft genome sequence of *Camellia sinensis* var. *sinensis* highlighted the impact of whole-genome duplications and subsequent paralogous duplications on the amplification of secondary metabolite genes, which are crucial for tea quality (Wei et al., 2018). These studies provide a theoretical basis for future research to explore the genetic and epigenetic factors underpinning the regulation of gene expression in tea.

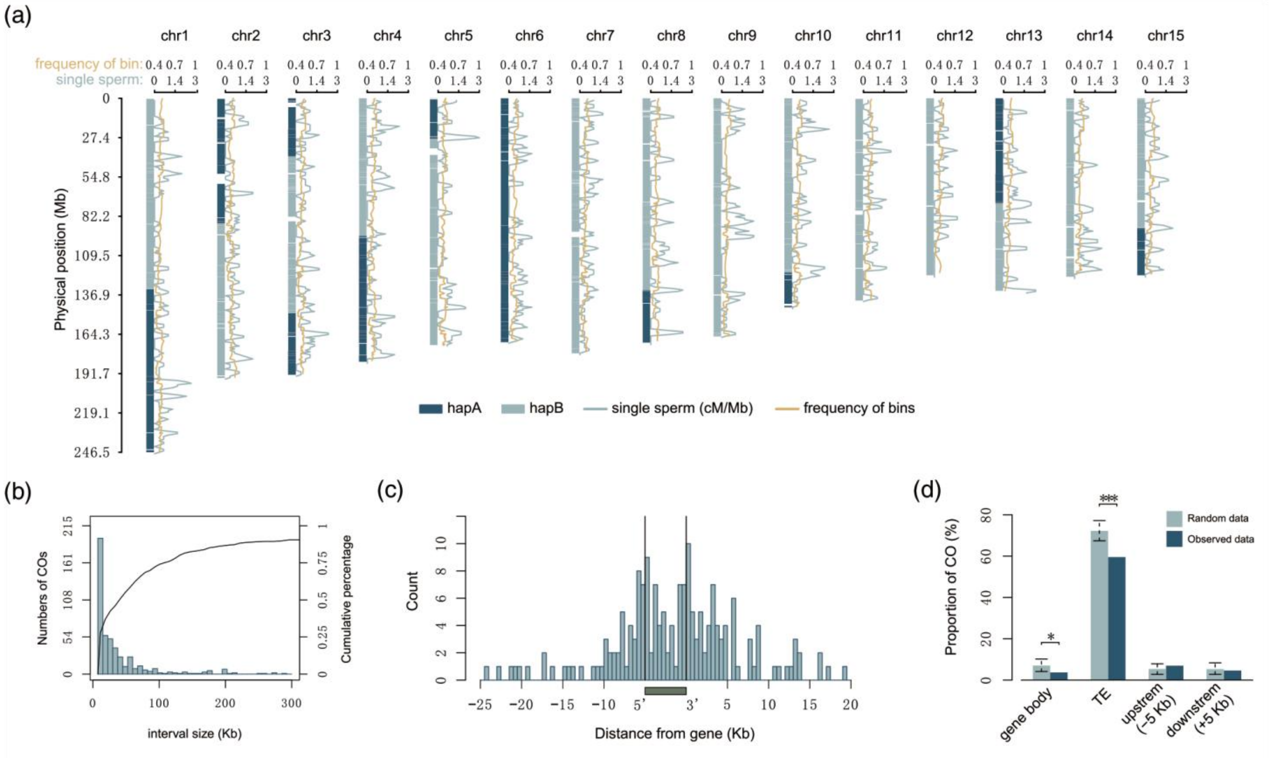

Figure 2 CO patterns detected by single sperm sequencing of Fudingdabai (Adopted from Zhang et al., 2020a)

Image caption: Bin map of single sperm cell ID "QSC-1", showing allelic frequencies in different segments; b: Distribution of CO resolution, with the histogram representing the number of COs and the black curve indicating the cumulative percentage of CO interval size; c: CO distribution at the gene scale, with dark green boxes representing the normalized gene range; d: Association between COs and genomic features in the tea plant genome, showing that most COs occur in transposable element regions (59.53%) and only 3.72% of COs are located within gene bodies. These results reveal the distribution patterns and influencing factors of COs in tea plants, providing important data for further genetic and breeding research in tea plants (Adapted from Zhang et al., 2020a)

5 Comparative Genomics

5.1 Benefits of comparative studies

Comparative genomics offers significant benefits in understanding the genetic and evolutionary relationships among different species. By comparing the genome of the tea plant with the genomes of other species, researchers can identify conserved genomic regions and lineage-specific innovations, elucidating the evolutionary relationships and differentiation patterns among different tea species (e.g., *Camellia sinensis* var. *sinensis* and *Camellia sinensis* var. *assamica*). Comparative analysis of genomic sequences reveals the genetic variations underlying various agronomic traits, such as leaf morphology, biochemical composition (e.g., catechin profiles), and environmental adaptability. For instance, the high-quality reference genome of *Camellia sinensis* var. *sinensis* reveals the critical role of long terminal repeat retrotransposons (LTR-RTs) in genome expansion and gene transcription diversification. Notably, genes associated with tea aroma and stress resistance have undergone significant amplification through recent tandem duplications, forming gene clusters that are crucial for the plant's adaptability and quality traits (Xia et al., 2020a).

Additionally, comparative studies can help trace the evolutionary history and domestication processes of tea plants, as evidenced by the phylogenetic analyses of diverse tea plant accessions. Through the high-quality assembly of the Camellia sinensis var. sinensis genome and the resequencing of 81 tea samples from different sources, researchers have identified key genes involved in the evolution and adaptation of the tea genome. Phylogenetic analysis supports the southwest origin of tea plants and reveals the historical spread of cultivated tea in China (Xia et al., 2020a).

5.2 Comparative genomic techniques

Several advanced techniques are employed in comparative genomics to analyze and compare the genomes of tea plants. High-throughput sequencing technologies, such as Illumina and PacBio, have been instrumental in generating high-quality genome assemblies of tea plants (Wei et al., 2018). These technologies enable the identification of gene family expansions, whole-genome duplications, and other genomic variations that contribute to the unique characteristics of tea plants. For example, the draft genome sequence of *Camellia sinensis* var. *sinensis* has facilitated the analysis of gene family evolution and the identification of key genes involved in the biosynthesis of important tea metabolites (Wei et al., 2018).

Using PacBio SMRT technology for high-quality full-length transcriptome sequencing of tea tree roots and young shoots has revealed the characteristics of metabolites such as catechins, theanine, and caffeine in different tissues. These studies provide a foundation for further metabolomics and gene regulation research in tea trees (Zhang et al., 2021b). Additionally, single sperm sequencing has been used to phase the genome of tea plants, providing high-resolution genetic and recombination maps that reveal crossover patterns and genetic relatedness among tea accessions (Zhang et al., 2020a).

5.3 Implications of comparative results

The results of comparative genomic studies have profound implications for tea plant breeding and functional genomics research. By identifying genes associated with desirable traits, such as tea quality and stress resistance, researchers can develop molecular markers for breeding programs aimed at improving these traits in tea plants (Xia et al., 2020b). Furthermore, understanding the genetic basis of tea plant adaptation and domestication can inform conservation strategies and the utilization of tea germplasm resources (Xia et al., 2020a). The insights gained from comparative genomics also contribute to our broader understanding of genome evolution in flowering plants, as demonstrated by the identification of whole-genome duplications and subsequent gene family expansions in the tea plant genome (Wei et al., 2018). These findings lay the foundation for future research aimed at unlocking the full potential of the tea genome for both scientific and agricultural advancements.

6 Epigenetics and Regulation

6.1 Overview of epigenetic modifications

Epigenetic modifications encompass heritable changes in gene expression that do not involve alterations in the DNA sequence itself. These modifications include DNA methylation, histone modification, and RNA-associated silencing. DNA methylation typically occurs at cytosine residues in the context of CpG dinucleotides and is a key mechanism for regulating gene expression and maintaining genome stability (Yuan, 2020; Zhao et al., 2020). Histone modifications, such as acetylation, methylation, and phosphorylation, alter chromatin structure and thereby influence gene accessibility and transcriptional activity (Jain et al., 2021). The advent of high-throughput sequencing technologies has significantly advanced our ability to map these modifications across the genome, providing insights into their roles in various biological processes (Yuan, 2020; Zhao et al., 2020).

In tea plants, epigenetic modifications dynamically respond to environmental cues, developmental stages, and stress conditions, modulating phenotypic plasticity and adaptation (Xia et al., 2020b). Understanding the epigenetic landscape of the tea genome is essential for unraveling its regulatory networks and harnessing epigenetic variation for crop improvement and sustainable agriculture.

6.2 Epigenetic regulation in tea plants

In tea plants (*Camellia sinensis*), epigenetic regulation plays a crucial role in controlling gene expression and contributing to phenotypic diversity. Recent studies have highlighted the importance of DNA methylation and histone modifications in regulating genes associated with tea quality, stress responses, and developmental processes (Xia et al., 2020b). For instance, the phased genome sequencing of the elite tea cultivar "Fudingdabai" has revealed allele-specific expression patterns and differential expression levels between haplotypes, suggesting a significant role for epigenetic mechanisms in tea plant regulation (Zhang et al., 2020a). The identification of epigenetic markers and their association with desirable traits can facilitate the development of improved tea cultivars through epibreeding (Jain et al., 2021).

Furthermore, small RNA-mediated epigenetic pathways, such as microRNAs (miRNAs) and small interfering RNAs (siRNAs), orchestrate gene silencing and post-transcriptional regulation of target genes involved in secondary metabolism and defense responses (Lee and Carroll, 2018). The interplay between epigenetic modifications and environmental factors, such as temperature fluctuations and nutrient availability, underscores their role in phenotypic variability and adaptive plasticity across diverse tea cultivars and growing conditions.

6.3 Future directions in epigenetic research

Future research in tea plant epigenetics should focus on several key areas. Comprehensive mapping of the tea epigenome at single-base resolution will provide a deeper understanding of the epigenetic landscape and its impact on gene regulation (Yuan, 2020; Zhao et al., 2020). Investigating the stability and heritability of epigenetic modifications across generations will be essential for harnessing epigenetic variation in breeding programs (Jain et al., 2021). Integrating epigenomic data with transcriptomic and metabolomic profiles will help elucidate the complex regulatory networks underlying important agronomic traits (Xia et al., 2020b). Furthermore, developing advanced tools and methodologies for precise epigenetic editing will open new avenues for crop improvement and functional genomic studies in tea plants (Yuan, 2020; Zhao et al., 2020; Jain et al., 2021). By advancing our understanding of epigenetic regulation in tea plants, we can unlock new potentials for enhancing tea quality, stress tolerance, and overall crop performance, ultimately benefiting both producers and consumers.

7 Case Study: Genome Studies in Select Tea Varieties

7.1 Selection of tea varieties for genomic analysis

The selection of tea varieties for genomic analysis is crucial to understanding the genetic diversity and evolutionary history of tea plants. Various studies have focused on different tea cultivars to uncover the genetic basis of important traits such as quality, drought tolerance, and disease resistance. For instance, a haplotype-resolved genome assembly of the Oolong tea cultivar, Tieguanyin, was conducted to explore allele-specific expression and the evolutionary history of *Camellia sinensis* (Zhang et al., 2021a). The study shows that through the analysis of allele-specific expression, a potential mechanism for coping with mutation load has been revealed in tea trees during long-term asexual reproduction (Zhang et al., 2021a). By conducting a population genomic analysis of 190 *Camellia* samples, the population history of *Camellia sinensis* var. *sinensis* (CSS) and *Camellia sinensis* var. *assamica* (CSA) was uncovered, along with the independent evolutionary histories and parallel domestication of the two major cultivated varieties (Figure 3). Additionally, the study highlights the role of gene selection in the flavor characteristics of tea trees and the green revolution of the tea industry, providing crucial genetic resources and molecular insights for gene editing and trait improvement in tea trees.

Another study focused on genomic selection in the breeding of black tea (*Camellia sinensis*). The research evaluated genome-enabled prediction models for tea quality and drought tolerance traits using machine learning techniques, including Extreme Learning Machine, Support Vector Machines, and Principal Component Analysis (Koech et al., 2019). The results showed that models combining QTLs (Quantitative Trait Loci), annotated proteins, and KEGG pathways had better predictive ability for catechins, astringency, brightness, briskness, and color of tea. The application of these models is expected to accelerate the tea breeding process, improve selection efficiency, and reduce costs and time (Koech et al., 2019).

7.2 Findings and implications

The findings from these genomic studies have significant implications for tea breeding and cultivation. The haplotype-resolved genome assembly of Tieguanyin revealed independent evolutionary histories and parallel domestication in two widely cultivated varieties, var. sinensis and var. assamica. It also uncovered extensive intra- and interspecific introgressions contributing to genetic diversity in modern cultivars (Zhang et al., 2021). These insights are crucial for understanding the genetic basis of flavor characteristics and other desirable traits in tea plants.

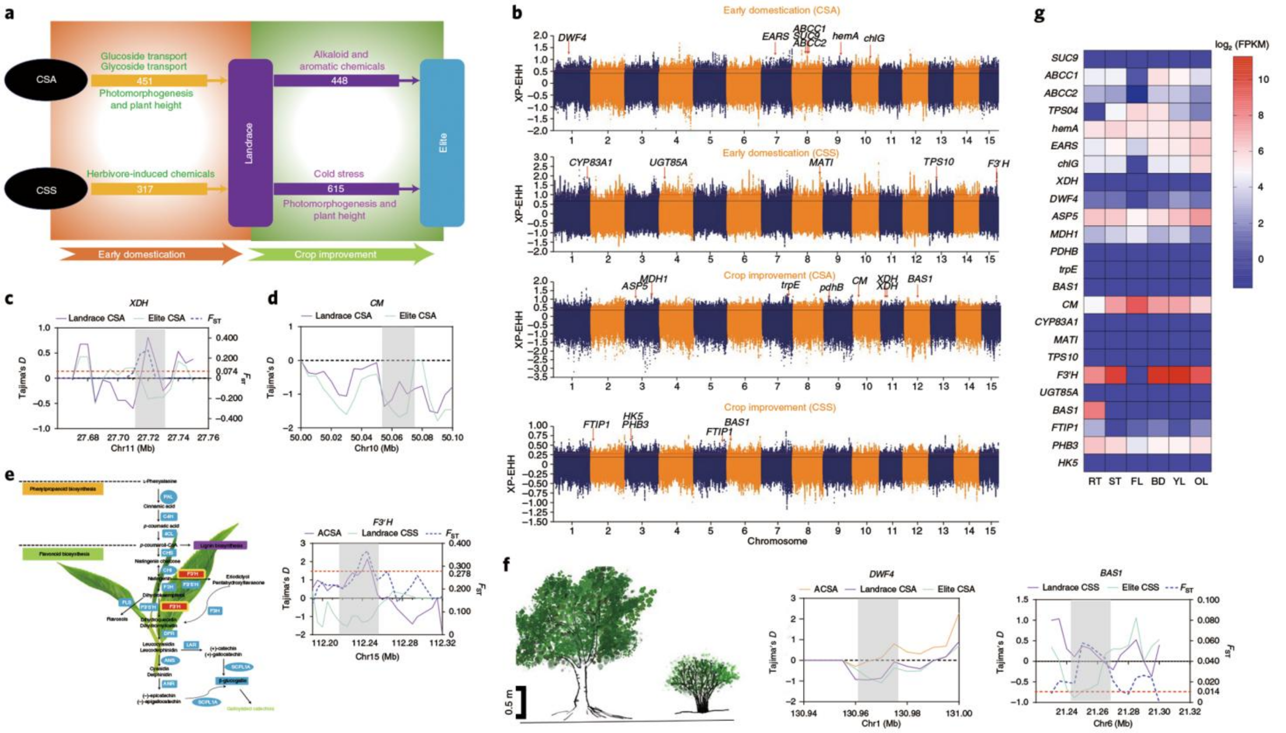

Figure 3 Signatures of artificial selection and evidence of parallel domestication in CSA and CSS (Adopted from Zhang et al., 2021a)

Image caption: a: The roadmap of parallel domestication in CSA and CSS. During early domestication, CSA mainly involved genes related to glucoside transport, photomorphogenesis, and plant height, while CSS involved chemicals for defense against insects. In the improvement process, CSA focused on the metabolism of alkaloids and aromatic compounds, while CSS mainly concentrated on genes related to cold stress and photomorphogenesis and plant height; b: Selective sweep signals identified based on XP-EHH (cross-population extended haplotype homozygosity). These signals indicate multiple selected gene regions in CSA and CSS; c: Artificial selection signals of the *XDH* gene (xanthine dehydrogenase-oxidase); d: Artificial selection signals of the *CM* gene (chorismate mutase), involved in the biosynthesis of aromatic amino acids. During early domestication of CSA varieties, the *CM* gene shows significant artificial selection signals; e: Strong artificial selection signals of the *F3'5'H* gene in catechin biosynthesis, showing high FST scores and significantly low Tajima's D values in CSS landraces; f: Artificial selection signals of BAS1 and *DWF4* genes related to plant height. The reduced plant height in cultivated CSA and CSS varieties is mainly associated with the selection of these genes; g: RNA-seq expression analysis results of artificially selected genes in different tissues, further supporting the potential functions of these genes in various tissues. The results reveal parallel evolutionary pathways during the domestication of CSA and CSS, and through the selection of these genes, tea plant varieties have been improved and adapted (Adapted from Zhang et al., 2021a)

In the study on genomic selection for black tea, the best prediction models were identified for various quality traits, such as catechin, astringency, brightness, briskness, and color. The use of putative QTLs, annotated proteins, and KEGG pathways in the prediction models showed robustness and usefulness in predicting phenotypes (Koech et al., 2019). This approach opens up new avenues for future applications of genomic selection in tea breeding, potentially accelerating the development of high-quality tea cultivars.

Overall, the advances in high-quality sequencing and annotation of tea genomes have provided a wealth of genetic and molecular information. These studies not only enhance our understanding of the genetic basis of important traits in tea plants but also provide a reference for developing improved tea varieties through genomic selection and gene editing technologies.

8 Applications in Agriculture and Breeding

8.1 Genomic applications in tea breeding

Recent advancements in tea plant genomics have significantly enhanced the breeding programs for tea plants. The integration of genomic predictions (GPs) and genome-wide association studies (GWASs) has shown potential in improving the genetic breeding of tea quality-related metabolites. For instance, the use of genome-wide single nucleotide polymorphisms (SNPs) detected from restriction site-associated DNA sequencing has facilitated the identification of candidate genes for key metabolites such as catechins and caffeine, thereby contributing to genomics-assisted tea breeding (Yamashita et al., 2020). Additionally, the high-quality genome assembly of Camellia sinensis var. sinensis has provided insights into the evolution of the tea genome and the biosynthesis of key tea metabolites, providing references for future research aimed at improving tea quality and diversity within tea germplasm (Wei et al., 2018).

8.2 Genetic modifications and their acceptance

The application of genetic modifications in tea plants, such as CRISPR-based genome editing, has the potential to accelerate breeding programs by enabling precise modifications of the genome. This approach has been exemplified in other crops like wheat, where genome editing has been used to modify traits such as flowering time and stress resistance (Appels et al., 2018). However, the acceptance of genetically modified tea plants may face challenges due to consumer perceptions and regulatory hurdles. It is crucial to engage with stakeholders, including consumers, policymakers, and industry players, to address concerns and highlight the benefits of genetic modifications in enhancing tea quality and sustainability.

8.3 Future potentials in agricultural applications

The future of tea plant breeding lies in the continued application of high-quality sequencing and annotation technologies. The reference genome of the tea plant and the resequencing of diverse accessions have provided valuable resources for understanding genome evolution and adaptation, which can be leveraged to develop improved tea varieties with enhanced quality and stress resistance (Xia et al., 2020a). Moreover, the phased genome based on single sperm sequencing has revealed complex relatedness and genetic signatures among tea accessions, offering insights into the regulation of gene expression and the evolution of tea plants (Zhang et al., 2020a). These genomic resources will play a pivotal role in future breeding efforts, enabling the development of tea varieties that are better adapted to changing environmental conditions and consumer preferences.

The advancements in tea plant genomics have opened new avenues for improving tea breeding programs. By leveraging genomic predictions, genome-wide association studies, and genetic modifications, researchers can accelerate the development of high-quality tea varieties. The continued application of high-quality sequencing and annotation techniques will further enhance our understanding of the tea plant genome, aiding in future innovations in tea agriculture and breeding.

9 Concluding Remarks

Recent advancements in high-quality sequencing and annotation have significantly enhanced our understanding of the tea plant genome. The sequencing of the *Camellia sinensis* var. *sinensis* genome has provided insights into the evolution of tea quality traits and the molecular mechanisms underlying these traits. The identification of specific gene families responsible for the biosynthesis of key metabolites such as catechins, theanine, and caffeine has been a major breakthrough. Additionally, the reference genome and resequencing of diverse tea plant accessions have shed light on the genome evolution and adaptation of tea plants, revealing the critical roles of LTR retrotransposons in genome size expansion and gene diversification. These studies have laid a solid foundation for future research aimed at improving tea quality and stress resistance through targeted breeding programs.

Despite the significant progress, several challenges remain in the field of tea genomics. One of the primary challenges is the complexity of the tea plant genome, which is characterized by a high percentage of repetitive sequences and multiple rounds of whole-genome duplications. This complexity makes it difficult to achieve complete and accurate genome assemblies. The functional annotation of genes, particularly those involved in secondary metabolite pathways, remains incomplete. There is also a need for more comprehensive studies on the epigenetics and noncoding RNAs in tea plants to fully understand their roles in gene regulation and trait expression. Furthermore, the genetic diversity within tea germplasm is vast, and more efforts are needed to explore and utilize this diversity for breeding purposes.

The future of tea genomics research holds great promise. Advances in sequencing technologies and bioinformatics tools will likely lead to more complete and accurate genome assemblies, facilitating deeper insights into the genetic basis of tea quality and stress resistance. Functional genomic studies, including gene editing and transcriptome analysis, will be crucial for identifying and manipulating key genes to enhance desirable traits in tea plants. Additionally, integrating genomics with other omics approaches, such as metabolomics and proteomics, will provide a more holistic understanding of the molecular mechanisms governing tea plant biology. The development of high-throughput phenotyping platforms and the application of machine learning algorithms will further accelerate the breeding of improved tea varieties with enhanced quality and resilience to environmental stresses. Ultimately, these advancements will contribute to the sustainable production of high-quality tea, benefiting both producers and consumers worldwide.

Acknowledgments

The HortHerb Publisher appreciate the feedback from two anonymous peer reviewers on the manuscript of this study.

Funding

This work was supported by the National Natural Science Foundation of China [No. 32160077] and the Guizhou Academy of Agricultural Sciences Talent Special Project [2022-02 and 2023-02].

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

Appels R., Eversole K., Stein N., et al., 2018, Shifting the limits in wheat research and breeding using a fully annotated reference genome, Science, 361(6403): eaar7191.
https://doi.org/10.1126/science.aar7191
PMid:30115783

Bansal G., Narta K., and Teltumbade M., 2018, Next-generation sequencing: technology, advancements, and applications, Bioinformatics: Sequences, Structures, Phylogeny, pp. 15-46.
https://doi.org/10.1007/978-981-13-1562-6_2
PMCid:PMC5750229

Chin C., Peluso P., Sedlazeck F., Nattestad M., Concepcion G., Clum A., Dunn C., O'Malley R., Figueroa-Balderas R., Morales-Cruz A., Cramer G., Delledonne M., Luo C., Ecker J., Cantu D., Rank D., and Schatz M., 2016, Phased diploid genome assembly with single-molecule real-time sequencing, Nature Methods, 13(12): 1050-1054.
https://doi.org/10.1038/nmeth.4035
PMid:27749838 PMCid:PMC5503144

Hazra A., Kumar R., Sengupta C., and Das S., 2020, Genome-wide SNP discovery from Darjeeling tea cultivars—their functional impacts and application toward population structure and trait associations, Genomics, 113(1): 66-78.
https://doi.org/10.1016/j.ygeno.2020.11.028
PMid:33276009

Jain N., Taak Y., Choudhary R., Yadav S., Saini N., Vasudev S., and Yadava D., 2021, Advances and prospects of epigenetics in plants, Epigenetics and Metabolomics, pp. 421-444.
https://doi.org/10.1016/b978-0-323-85652-2.00013-0

Koech R.K., Malebe P.M., Nyarukowa C., Mose R., Kamunya S.M., Loots T., and Apostolides Z., 2020, Genome-enabled prediction models for black tea (*Camellia sinensis*) quality and drought tolerance traits, Plant Breeding, 139(5): 1003-1015.
https://doi.org/10.1101/850792

Kong W., Jiang M., Wang Y., Chen S., Zhang S., Lei W., Chai K., Wang P., Liu R., and Zhang X., 2022, Pan-transcriptome assembly combined with multiple association analysis provides new insights into the regulatory network of specialized metabolites in the tea plant *Camellia sinensis*, Horticulture Research, 9: uhac100.
https://doi.org/10.1093/hr/uhac100

Kronenberg Z.N., Hall R.J., Hiendleder S., Smith T.P., Sullivan S.T., Williams J.L., and Kingan S.B., 2018, FALCON-Phase: integrating PacBio and Hi-C data for phased diploid genomes, BioRxiv, 327064.
https://doi.org/10.1101/327064

Lee C.H., and Carroll B.J., 2018, Evolution and diversification of small RNA pathways in flowering plants, Plant and Cell Physiology, 59(11): 2169-2187.
https://doi.org/10.1093/pcp/pcy167
PMCid:PMC6454791

Li F.D., Tong W., Xia E.H., and Wei C.L., 2019, Optimized sequencing depth and de novo assembler for deeply reconstructing the transcriptome of the tea plant, an economically important plant species, BMC Bioinformatics, 20: 1-11.
https://doi.org/10.1186/s12859-019-3166-x
PMid:31694521 PMCid:PMC6836513

Mgwatyu Y., Cornelissen S., van Heusden P., Stander A., Ranketse M., and Hesse U., 2022, Establishing MinION sequencing and genome assembly procedures for the analysis of the rooibos (*Aspalathus linearis*) genome, Plants, 11(16): 2156.
https://doi.org/10.3390/plants11162156
PMid:36015459 PMCid:PMC9416007

Ou S., Su W., Liao Y., Chougule K., Agda J., Hellinga A., Lugo C., Elliott T., Ware D., Peterson T., Jiang N., Hirsch C., and Hufford M., 2019, Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline, Genome Biology, 20: 1-18.
https://doi.org/10.1186/s13059-019-1905-y
PMid:31843001 PMCid:PMC6913007

Shi C., Yang H., Wei C., Yu O., Zhang Z., Jiang C., Sun J., Li Y., Chen Q., Xia T., and Wan X., 2011, Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds, BMC Genomics, 12: 1-19.
https://doi.org/10.1186/1471-2164-12-131

Speranskaya A.S., Khafizov K., Ayginin A.A., Krinitsina A.A., Omelchenko D.O., Nilova M.V., Severova E., Samokhina E., Shipulin G., and Logacheva M., 2018, Comparative analysis of Illumina and Ion Torrent high-throughput sequencing platforms for identification of plant components in herbal teas, Food Control, 93: 315-324.
https://doi.org/10.1016/J.FOODCONT.2018.04.040

Wang F., Chen Z., Pei H., Guo Z., Wen D., Liu R., and Song B., 2021, Transcriptome profiling analysis of tea plant (*Camellia sinensis*) using Oxford Nanopore long-read RNA-Seq technology, Gene, 769: 145247.
https://doi.org/10.1016/j.gene.2020.145247
PMid:33096183

Wei C., Yang H., Wang S., Zhao J., Liu C., Gao L., Xia E., Lu Y., Tai Y., She G., Sun J., Cao H., Tong W., Gao Q., Li Y., Deng W., Jiang X., Wang W., Chen Q., Zhang S., Li H., Wu J., Wang P., Li P., Shi C., Zheng F., Jian J., Huang B., Shan D., Shi M., Fang C., Yue Y., Li F., Li D., Wei S., Han B., Jiang C., Yin Y., Xia T., Zhang Z., Bennetzen J., Zhao S., and Wan X., 2018, Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality, Proceedings of the National Academy of Sciences, 115(18): E4151-E4158.
https://doi.org/10.1073/pnas.1719622115
PMid:29678829 PMCid:PMC5939082

Xia E., Tong W., Hou Y., An Y., Chen L., Wu Q., Liu Y., Yu J., Li F., Li R., Li P., Zhao H., Ge R., Huang J., Mallano A., Zhang Y., Liu S., Deng W., Song C., Zhang Z., Zhao J., Wei S., Zhang Z., Xia T., Wei C., and Wan X., 2020a, The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation, Molecular Plant, 13(7): 1013-1026.
https://doi.org/10.1016/j.molp.2020.04.010
PMid:32353625

Xia E., Tong W., Wu Q., Wei S., Zhao J., Zhang Z., Wei C., and Wan X., 2020b, Tea plant genomics: achievements, challenges and perspectives, Horticulture Research, 7.
https://doi.org/10.1038/s41438-019-0225-4

Xia E., Zhang H., Sheng J., Li K., Zhang Q., Kim C., Zhang Y., Liu Y., Zhu T., Li W., Huang H., Tong Y., Nan H., Shi C., Shi C., Jiang J., Mao S., Jiao J., Zhang D., Zhao Y., Zhao Y., Zhang L., Liu Y., Liu B., Yu Y., Shao S., Ni D., Eichler E., and Gao L., 2017, The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis, Molecular Plant, 10(6): 866-877.
https://doi.org/10.1016/j.molp.2017.04.002

Yamashita H., Uchida T., Tanaka Y., Katai H., Nagano A.J., Morita A., and Ikka T., 2020, Genomic predictions and genome-wide association studies based on RAD-seq of quality-related metabolites for the genomics-assisted breeding of tea plants, Scientific Reports, 10(1): 17480.
https://doi.org/10.1038/s41598-020-74623-7

Yuan B.F., 2019, Assessment of DNA epigenetic modifications, Chemical Research in Toxicology, 33(3): 695-708.
https://doi.org/10.1021/acs.chemrestox.9b00372
PMid:31690070

Zhang W., Luo C., Scossa F., Zhang Q., Usadel B., Fernie A., Mei H., and Wen W., 2020a, A phased genome based on single sperm sequencing reveals crossover pattern and complex relatedness in tea plants, The Plant Journal, 105(1): 197-208.
https://doi.org/10.1111/tpj.15051
PMid:33118252

Zhang W., Zhang Y., Qiu H., Guo Y., Wan H., Zhang X., Scossa F., Alseekh S., Zhang Q., Wang P., Xu L., Schmidt M., Jia X., Li D., Zhu A., Guo F., Chen W., Ni D., Usadel B., Fernie A., and Wen W., 2020b, Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties, Nature Communications, 11(1): 3719.
https://doi.org/10.1038/s41467-020-17498-6

Zhang X., Chen S., Shi L., Gong D., Zhang S., Zhao Q., Zhan D., Vasseur L., Wang Y., Yu J., Liao Z., Xu X., Qi R., Wang W., Ma Y., Wang P., Ye N., Ma D., Shi Y., Wang H., Ma X., Kong X., Lin J., Wei L., Ma Y., Li R., Hu G., He H., Zhang L., Ming R., Wang G., Tang H., and You M., 2021a, Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*, Nature Genetics, 53(8): 1250-1259.
https://doi.org/10.1038/s41588-021-00895-y
PMid:34267370 PMCid:PMC8346365

Zhang Y., Li P., She G., Xu Y., Peng A., Wan X., and Zhao J., 2021, Molecular basis of the distinct metabolic features in shoot tips and roots of tea plants (*Camellia sinensis*): characterization of MYB regulator for root theanine synthesis, Journal of Agricultural and Food Chemistry, 69(11): 3415-3429.
https://doi.org/10.1021/acs.jafc.0c07572
PMid:33719427

Zhao L., Song J., Liu Y., Song C., and Yi C., 2020, Mapping the epigenetic modifications of DNA and RNA, Protein & Cell, 11(11): 792-808.
https://doi.org/10.1007/s13238-020-00733-7
PMid:32440736 PMCid:PMC7647981

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.