Tree Genetics and Molecular Breeding 2024, Vol.14, No.2, 69-80
http://genbreedpublisher.com/index.php/tgmb

Feature Review **Open Access**

Poplar Genome Analysis and Functional Gene Annotation and Verification

Xiaojing Yang, Yuxin Zhu
Modern Agricultural Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China
Corresponding email: yuxin.zhu@cuixi.org

Tree Genetics and Molecular Breeding, 2024, Vol.14, No.2 doi: 10.5376/tgmb.2024.14.0008
Received: 13 Feb., 2024
Accepted: 15 Mar., 2024
Published: 23 Mar., 2024

Preferred citation for this article:
Yang X.J., and Zhu Y.X., 2024, Poplar genome analysis and functional gene annotation and verification, Tree Genetics and Molecular Breeding, 14(2): 69-80 (doi: 10.5376/tgmb.2024.14.0008)

Abstract This study explores poplar gene annotation techniques, including methods for annotating functional genes, challenges in annotation accuracy, and advances in computational tools for genome annotation. Through specific case studies, key functional genes in poplar are analyzed, and their functions are experimentally verified, with comparative analyses conducted against gene function studies in model plants. Functional genomics plays a critical role in poplar research, involving the use of gene knockout and overexpression studies to verify gene functions, and the integration of transcriptomic and proteomic data to support gene annotation. Insights from the poplar genome are also applied in biotechnological and environmental management, such as improving wood quality and growth rates, enhancing stress resistance and phytoremediation capabilities, and developing poplar breeding programs. Emerging genomic technologies such as CRISPR and other genome editing tools are revolutionizing approaches to studying the poplar genome. This research aims to advance poplar genomics studies, providing scientific support for sustainable forestry.

Keywords Poplar; Gene annotation; Functional genomics; Biotechnology; Sustainable forestry

1 Introduction

Poplar trees (*Populus* L.), belonging to the genus *Populus*, are of immense significance in both forestry and biotechnological applications. They are widely cultivated for their rapid growth, ease of vegetative propagation, and small genome size, making them ideal for genetic studies and biotechnological interventions (Confalonieri et al., 2003; Brunner et al., 2004). Poplars provide a range of ecological services, including carbon sequestration, bioremediation, and habitat provision, which are crucial for maintaining ecological balance. Additionally, they are valuable for the forest products industry, contributing to the production of timber, pulp, and paper, and are increasingly recognized for their potential in renewable energy production (Chen et al., 2023). The ability to genetically engineer poplars to enhance traits such as pest resistance, herbicide tolerance, and wood quality further underscores their importance in forestry and biotechnology.

The sequencing of the poplar genome has marked a significant milestone in plant genomics, providing a comprehensive resource for understanding the genetic basis of various traits in trees (Brunner et al., 2004; Tuskan et al., 2004). The U.S. Department of Energy's initiative to sequence the *Populus trichocarpa* genome has enabled researchers to compare the genetic makeup of perennial trees with annual plants, offering insights into tree-specific genetic programs. The availability of extensive expressed sequence tags (ESTs) and full-length cDNA sequences has facilitated the annotation of the poplar genome, allowing for the identification of genes involved in critical processes such as cellulose biosynthesis, secondary xylem formation, and defense against biotic stress (Joshi et al., 2004; Ralph et al., 2008). These genomic resources have also paved the way for the development of functional genomics tools, such as cDNA microarrays, to study gene expression in response to environmental stimuli.

This study utilizes extensive genomic data to analyze the genetic mechanisms behind key traits of poplar trees such as growth, development, and stress responses. It also explores the potential applications of these findings in improving poplar genotypes for use in forestry and biotechnology, to address challenges such as climate change and sustainable forest management. The research aims to advance poplar breeding programs and the sustainable management of forest resources.

2 Overview of Poplar Genome Structure

2.1 Characteristics and complexity of the poplar genome

The poplar genome is notable for its modest size and extensive genetic diversity, making it an ideal model for tree molecular biology and biotechnology. The genome of *Populus trichocarpa*, for instance, has been sequenced to an approximately 6x depth, revealing a genome size of around 520 Mbp (Brunner et al., 2004). This genome size is relatively small compared to other tree species, facilitating easier manipulation and study. The poplar genome also exhibits high levels of structural variation (SV), including insertions, deletions, and copy number variations (CNVs), which play significant roles in genome evolution and adaptation (Pinosio et al., 2016). These SVs are often located in low-gene density regions and are associated with transposable elements, indicating a dynamic genome structure.

2.2 Major milestones in poplar genome sequencing

The sequencing of the poplar genome has been a collaborative international effort, spearheaded by the U.S. Department of Energy. The annotated whole genome sequence of *Populus trichocarpa* was released to the public in early 2004, marking a significant milestone in tree genomics (Tuskan et al., 2004). This project provided the first opportunity to compare the genome of a perennial tree with that of annual plants, offering insights into tree-specific genetic traits such as dormancy and long-term host-pest interactions. Additionally, the assembly of the Populus alba genome, which is highly collinear with P. trichocarpa, further expanded the genomic resources available for poplar research (Ma et al., 2018). The development of comprehensive functional genomics resources, including cDNA libraries and microarray platforms, has also been crucial in advancing our understanding of poplar gene expression and regulation (Ralph et al., 2006; 2008).

2.3 Comparative genomics

Comparative genomics has revealed significant insights into the unique features of the poplar genome in relation to other tree species. For instance, the coding content of the poplar genome shows high similarity to that of Arabidopsis, an annual plant, suggesting that differences between these species are primarily due to gene regulation rather than gene content (Sterky et al., 2004). This similarity allows researchers to leverage the extensive functional genomic information available for Arabidopsis to study poplar 8. Furthermore, the expansion of certain gene families in poplar, such as those related to histone and auxin, highlights the evolutionary adaptations of poplar to its perennial lifestyle (Ma et al., 2018). Comparative analysis with other plant species has also identified unique multigene families in poplar, which may be related to its adaptation to environmental stresses (Park et al., 2005).

3 Gene Annotation Techniques in Poplar

3.1 Methods for annotating functional genes in the poplar genome

Gene annotation in poplar involves a combination of high-throughput sequencing, computational tools, and manual curation. One prominent method is the use of RNA-Seq data to identify and annotate carbohydrate-active enzymes (CAZymes) in the *Populus trichocarpa* genome. This approach allows for the identification of genes involved in the biosynthesis of cell wall polymers and other important metabolic processes by analyzing gene expression patterns across different tissues (Kumar et al., 2019). Another method involves the sequencing and annotation of large genomic sequences, such as a 95-kb stretch of *Populus deltoides*, which revealed disease resistance genes and transposable elements. Tools like ANNOVAR are also employed to annotate genetic variants by examining their functional consequences on genes and identifying variants in conserved regions (Wang et al., 2010).

3.2 Challenges in accurate gene annotation

Accurate gene annotation faces several challenges, including the high rate of misannotations due to the reliance on sequence homology and the propagation of errors through databases. Misannotations can be detected using methods that analyze genomic correlations to identify genes with unusually weak correlations in their assigned network positions (Hsiao et al., 2009). Another challenge is the integration of diverse types of functional genomic data, which requires sophisticated computational frameworks to combine data from different sources, such as microarray analyses and protein interaction data (Kemmeren et al., 2002; Troyanskaya et al., 2003). The complexity of the poplar genome, with its large number of paralogous genes and transposable elements, further complicates the annotation process (Lescot et al., 2004).

3.3 Advances in computational tools for genome annotation

Recent advances in computational tools have significantly improved the accuracy and efficiency of genome annotation. Tools like GenoCanyon use unsupervised statistical learning to integrate multiple computational and experimental annotations, predicting functional regions in the genome with high accuracy (Lu et al., 2015). The development of semi-automated methods, such as SamPler, allows for the optimization of annotation algorithms by selecting the best parameters based on manual curation of a sample set of genes (Cruz et al., 2019). Additionally, bioinformatics enrichment tools categorize and analyze large gene lists to identify biological processes relevant to the study, enhancing the functional annotation of genes (Huang et al., 2008). These advancements provide powerful resources for annotating the poplar genome and understanding its functional genomics.

4 Case Studies

4.1 Detailed analysis of key functional genes in poplar

The analysis of key functional genes in poplar has been significantly advanced through various genomic studies. For instance, a comprehensive genome-wide analysis of structural variation in three intercrossable poplar species (*Populus nigra*, *Populus deltoides*, and *Populus trichocarpa*) identified numerous insertions and deletions (INDELs) and copy number variations (CNVs) affecting thousands of genes. These structural variations were found to be associated with transposable elements and stress response genes, providing insights into the functional dynamics of the poplar genome (Figure 1) (Pinosio et al., 2016).

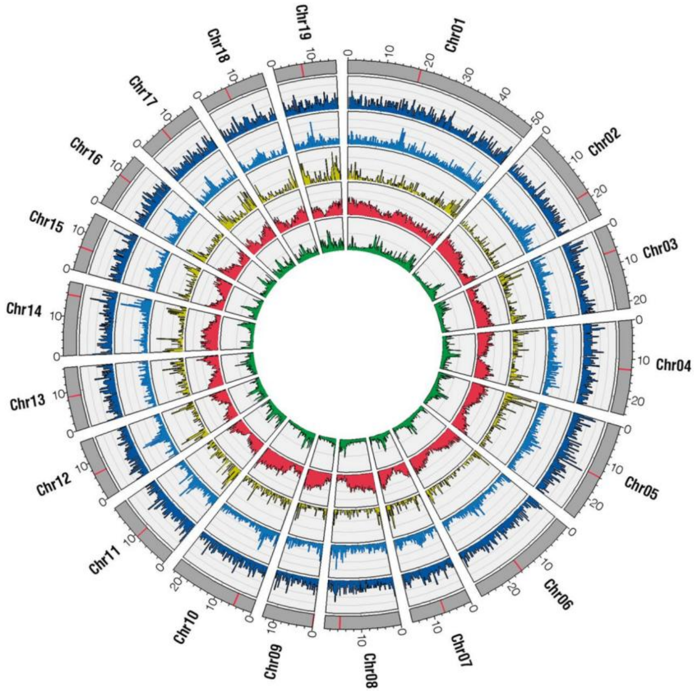

Figure 1 Genomic distribution of INDELs and genic CNVs (Adopted from Pinosio et al., 2016)

Image caption: The number of deletions (dark blue track, y axes range = 0-20), insertions (blue track, y axes range = 0-40), genic CNVs (yellow track, y axes range = 0-15), annotated genes (red track, y axes range = 0-80), and the repetitiveness of the genome (green track, y axes range = 0-10) are represented in windows of 250 kb along the 19 *Populus trichocarpa* chromosomes (outer gray bars). Predicted centromeric regions are highlighted in red (Adopted from Pinosio et al., 2016)

Pinosio et al. (2016) demonstrated structural variations in the poplar genome, which are predominantly concentrated in regions of low gene density and closely associated with transposon activity. Genes affected by structural variations exhibit lower expression levels than average, and these genes have higher dN/dS ratios, indicating they are under less selective pressure or are pseudogenes. These structural variations have significant impacts on the phenotypic diversity and genomic evolution of poplars. Transgenic technologies showcase tremendous potential in synthetic biology, improving industrial applications of wood such as biofuel production by optimizing lignin biosynthesis and cell wall structure.

Additionally, the PoplarGene network offers extensive functional annotations and interactions for approximately 70% of poplar genes, facilitating the understanding of biological processes such as wood development (Liu et al., 2016). Furthermore, the annotation of 19 841 full-length cDNA clones from *Populus nigra* has enriched the functional genomics resources, aiding in the accurate molecular annotation of poplar genes (Nanjo et al., 2007).

4.2 Success stories of gene verification through experimental approaches

Experimental approaches have played a crucial role in verifying gene functions in poplar. For example, the sequencing and annotation of a 95-kb genomic sequence from *Populus deltoides* revealed a cluster of disease resistance genes and novel transposable elements, which were experimentally validated to understand their roles in disease resistance (Lescot et al., 2004). Another success story involves the generation of a high-quality full-length cDNA collection for poplar, which was used to identify genes responding to insect feeding. This collection has been instrumental in reassessing gene predictions and identifying differentially expressed transcripts associated with defense mechanisms (Ralph et al., 2008). Additionally, the use of comparative genomics and functional genomic data from model microorganisms has enabled the prediction and experimental validation of functions for previously uncharacterized protein families, demonstrating the power of integrative approaches in gene function discovery (Gerdes et al., 2011).

4.3 Comparative analysis with gene function studies in model plants

Comparative analysis with model plants such as Arabidopsis thaliana has provided valuable insights into gene functions in poplar. The PLAZA 3.0 platform, which integrates comparative genomics data for 37 plant species, including poplar and Arabidopsis, allows for the transfer of functional annotations between species. This platform has facilitated the study of genome organization and gene functions across different plant species (Proost et al., 2014). Moreover, the re-annotation of carbohydrate-active enzyme (CAZyme) genes in *Populus trichocarpa*, based on RNA-Seq data, revealed similarities and differences in gene families between poplar and Arabidopsis, highlighting the evolutionary conservation and divergence of these enzymes (Kumar et al., 2019). The standardized genome-wide function prediction using the GOMAP pipeline has further enabled comparative functional genomics analyses, demonstrating that functional annotations across multiple species can retain sufficient biological signal to recover known phylogenetic relationships (Fattel et al., 2021).

5 Functional Genomics in Poplar Research

5.1 Role of functional genomics in unraveling gene functions

Functional genomics plays a crucial role in understanding the complex biological functions of genes within the poplar genome. By integrating various high-throughput techniques, researchers can systematically investigate gene functions and their interactions. For instance, the integration of transcriptomic and proteomic data has been shown to significantly enhance the accuracy of genome assembly and gene annotation, as demonstrated in studies involving *Anopheles stephensi* (Prasad et al., 2017). This approach can be similarly applied to poplar to identify and correct incomplete genome assemblies and to discover novel genes and gene functions.

5.2 Techniques for gene function verification

Gene function verification in poplar research often involves techniques such as gene knockout and overexpression studies. These methods allow researchers to observe the phenotypic consequences of gene disruption or overexpression, thereby elucidating gene functions. For example, the use of high-throughput sequencing and functional annotation tools like ANNOVAR can help identify functionally important genetic variants and their effects on gene expression (Wang et al., 2010). Additionally, the application of Bayesian frameworks to integrate heterogeneous data sources can improve the accuracy of gene function predictions, as seen in studies on *Saccharomyces cerevisiae*.

5.3 Integration of transcriptomic and proteomic data

The integration of transcriptomic and proteomic data is essential for accurate gene annotation in poplar research. This combined approach allows for a more comprehensive understanding of gene expression and protein function. For instance, integrating deep transcriptome and proteome data has been shown to correct genome assembly errors and improve gene model predictions, leading to the discovery of novel exons and protein-coding genes (Prasad et al., 2017). Furthermore, the use of bioinformatics enrichment tools can facilitate the functional analysis of large gene lists, enhancing the identification of relevant biological processes (Huang et al., 2008). By leveraging these integrated data sets, researchers can achieve more precise and reliable gene annotations in poplar.

6 Applications of Genomic Insights in Poplar

6.1 Biotechnological applications

Genomic insights have significantly advanced the biotechnological applications in poplar, particularly in improving wood quality and growth rates. By engineering lignin biosynthesis and modifying cell wall biogenesis, researchers have been able to enhance the properties of wood, making it more suitable for industrial applications (Tang et al., 2003). For example, the bicistronic expression of genes such as *PdGA20ox1* and *PtrMYB221* in transgenic poplars has resulted in increased biomass production and improved wood quality, demonstrating the potential of synthetic biology in this field (Figure 2) (Cho et al., 2018).

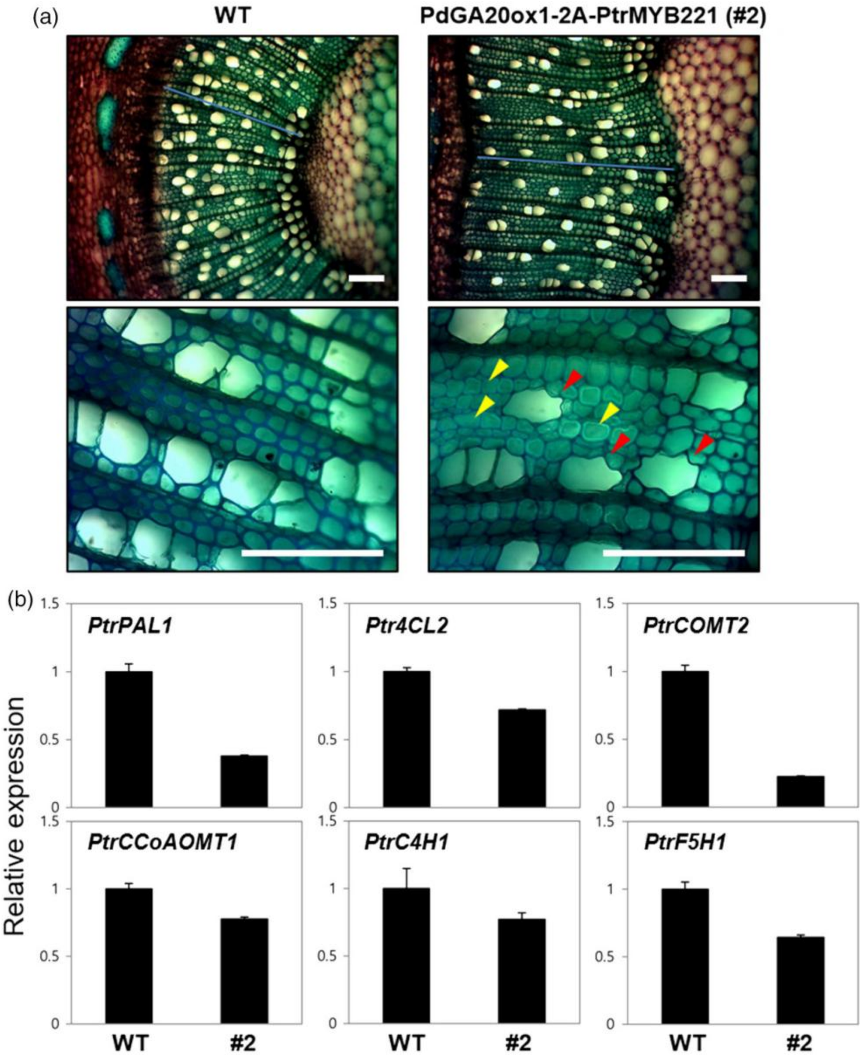

Figure 2 Transgenic poplars produce improved quality of woody biomass (Adopted from Cho et al., 2018)

Image caption: (a) Enhanced wood formation but reduced secondary wall thickening in DX15::PdGA20ox1-2A-PtrMYB221 transgenic poplars; (b) Expression of genes involved in the lignin biosynthetic pathway (Adopted from Cho et al., 2018)

Cho et al. (2018) found that the wood of genetically modified poplar trees not only exhibited a significant reduction in lignin content (by approximately 16%) but also an increase in cellulose content, thereby enhancing the overall quality of the wood. This transgenic technology demonstrates tremendous potential in the field of synthetic biology by optimizing lignin biosynthesis and cell wall structure, which can improve the industrial applications of wood, such as in the production of biofuels.

The sequencing of the poplar genome has facilitated the identification of key genes involved in wood formation and growth regulation. For instance, the development of the PoplarGene network has enabled researchers to prioritize genes related to wood development and other traits, thereby enhancing the understanding of biological processes underlying wood quality (Liu et al., 2016). Additionally, the discovery of single nucleotide polymorphisms (SNPs) in genes expressed in developing xylem has provided valuable resources for studying wood formation and secondary growth (Geraldes et al., 2011).

6.2 Environmental applications

Genomic insights have also been pivotal in enhancing the environmental applications of poplar, particularly in improving stress resistance and phytoremediation capabilities. The characterization of full-length enriched expressed sequence tags (ESTs) from stress-treated poplar leaves has provided a comprehensive understanding of the molecular responses to environmental stressors such as dehydration, salinity, and temperature extremes (Nanjo et al., 2004). This knowledge has facilitated the development of strategies to improve the environmental stress tolerance of poplar, making it a more resilient species in the face of climate change.

Moreover, genetic engineering has been employed to enhance the phytoremediation capabilities of poplar. For instance, transgenic poplar trees expressing the yeast cadmium factor 1 (*ScYCF1*) gene have shown increased tolerance to heavy metals and improved accumulation capacity, making them suitable for phytoremediation of contaminated soils (Shim et al., 2013). These transgenic plants exhibited enhanced growth and reduced toxicity symptoms in mine tailing soil, highlighting the potential of genetic engineering in environmental applications.

6.3 The Role of genomic information in poplar breeding programs

The integration of genomic information into poplar breeding programs has revolutionized the field, enabling more precise and efficient breeding strategies. The sequencing and annotation of the poplar genome have provided a wealth of genetic information that can be leveraged to identify and select desirable traits. For example, the identification of structural variations (SVs) and their association with stress resistance and pathogen defense has provided valuable markers for breeding programs aimed at enhancing these traits (Pinosio et al., 2016).

Additionally, the availability of comprehensive gene interaction networks, such as PoplarGene, has facilitated the identification of candidate genes for targeted breeding efforts (Liu et al., 2016). The use of genome editing technologies, such as CRISPR/Cas9, has further expedited the pace of poplar improvement programs by allowing precise modifications of specific genes involved in growth, development, and stress responses (Thakur et al., 2021). These advancements underscore the critical role of genomic information in modern poplar breeding programs, paving the way for the development of superior poplar varieties with enhanced traits.

7 Advances in Genomic Technologies

7.1 The impact of next-generation sequencing on poplar genomics

Next-generation sequencing (NGS) has significantly advanced the field of genomics by enabling rapid and comprehensive sequencing of entire genomes. This technology has been pivotal in the study of poplar genomics, allowing researchers to obtain detailed genetic information within a short time frame. NGS technologies have facilitated the identification of genetic variations and mutations, which are crucial for understanding the genetic basis of traits and diseases in poplar. The ability to sequence whole genomes or exomes has also accelerated the functional annotation of genes, providing insights into their roles and interactions (Schulze and Lammers, 2020). The continuous improvement of NGS technologies promises to further enhance our understanding of poplar genomics by providing more accurate and comprehensive data (Schulze and Lammers, 2020).

7.2 Emerging technologies

The advent of CRISPR/Cas9 and other genome editing tools has revolutionized the field of functional genomics. CRISPR/Cas9, in particular, has become a widely adopted tool due to its high efficiency, ease of use, and precision in editing specific genomic loci. This technology allows for the addition or removal of alleles, enabling the study of gene function and the development of genetically modified Poplar with desirable traits (Manghwar et al., 2019; Zhang et al., 2021). Other genome editing tools, such as zinc-finger nucleases (ZnF) and transcription activator-like effector nucleases (TALENs), also play significant roles in genome engineering, although CRISPR/Cas9 remains the most prominent due to its versatility and efficiency. Emerging CRISPR systems, including base editing and prime editing, offer even more precise and efficient genome editing capabilities, further expanding the potential applications in poplar genomics (Manghwar et al., 2019; Erwood et al., 2021).

7.3 Future outlook

Looking ahead, several emerging technologies are poised to further enhance genome analysis in Poplar. One such technology is CRISPR prime editing, which allows for precise and efficient introduction of genetic variants, addressing some of the limitations of traditional CRISPR/Cas9 editing (Erwood et al., 2021). Additionally, the development of high-throughput functional genomics tools, such as CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa), enables large-scale genetic screens to identify and characterize gene functions (Shalem et al., 2015; Ford et al., 2019; Zhang et al., 2021). These technologies, combined with advancements in single-cell sequencing and other high-resolution genomic techniques, will provide deeper insights into the complex regulatory networks and functional elements within the Poplar genome. As these technologies continue to evolve, they will undoubtedly play a crucial role in advancing our understanding and manipulation of Poplar genetics for both basic research and applied breeding programs (Hartenian and Doench, 2015; Hilton and Gersbach, 2015; Yan et al., 2015). By integrating these advanced genomic technologies, researchers can achieve a more comprehensive and functional understanding of the Poplar genome, paving the way for innovative applications in forestry, bioenergy, and environmental sustainability.

8 Integrating Genomic Data with Ecological and Physiological Studies

8.1 Ecological genomics

Ecological genomics aims to elucidate how genetic variations within poplar species influence their interactions with the environment. The sequencing of the poplar genome has provided a robust foundation for understanding these interactions. Poplars, being ecologically dominant, contribute significantly to carbon sequestration, nutrient cycling, and habitat formation (Brunner et al., 2004). The integration of genomic data with ecological studies has revealed insights into the genetic programs controlling ecological adaptation and environmental physiology (Brunner et al., 2004). For instance, the study of hybrid poplar interactions with forest tent caterpillars has highlighted the role of specific genes in plant defense mechanisms, which are crucial for ecological resilience (Ralph et al., 2006). Additionally, the genomic landscape of poplar has been mapped to understand local adaptation to environmental gradients, such as temperature, which is critical for predicting responses to climate change (Fitzpatrick and Keller, 2015).

8.2 Contributions of genomic studies to the physiology of poplar

Genomic studies have significantly advanced our understanding of the physiological processes in poplar. The availability of extensive expressed sequence tags (ESTs) and full-length cDNA libraries has facilitated the identification of genes involved in key physiological functions such as growth, development, and stress responses (Ralph et al., 2006). For example, the PoplarGene network provides a comprehensive resource for studying gene interactions related to wood development and other physiological traits. Moreover, the integration of gene expression data from various stress conditions has enabled the identification of genes that contribute to poplar's physiological adaptations to environmental stressors (Liu et al., 2021). These genomic resources are instrumental in dissecting the molecular mechanisms underlying poplar's physiological responses and improving its resilience to biotic and abiotic stresses.

8.3 Case Studies of genomic applications in ecological restoration

Genomic applications have shown great promise in ecological restoration efforts involving poplar species. The characterization of the poplar pan-genome has revealed structural variations that play a role in stress resistance and adaptation, which are essential for successful restoration projects (Pinosio et al.., 2016). Additionally, the use of genomic data to model the geographic distribution of genetic variation in response to environmental changes has provided valuable insights for selecting appropriate genotypes for restoration in different ecological contexts (Fitzpatrick and Keller, 2015). Case studies have demonstrated the utility of genomic tools in identifying and prioritizing genes for functional studies, which can inform breeding programs aimed at enhancing traits such as disease resistance and environmental tolerance (Ralph et al., 2008; Liu et al., 2016). These genomic applications are paving the way for more effective and sustainable ecological restoration strategies using poplar species.

9 Future Research Directions in Poplar Genomics

9.1 Unexplored areas and emerging fields in poplar genomic research

Despite significant advancements in poplar genomics, several areas remain underexplored. One such area is the comprehensive understanding of structural variations (SVs) and their impact on the poplar genome. Recent studies have highlighted the presence of numerous insertions and deletions (INDELs) and their association with transposable elements, suggesting a complex genomic architecture that warrants further investigation (Pinosio et al.., 2016). Additionally, the role of gene dosage and its effects on quantitative traits and heterosis in poplar is another emerging field. The development of a genome-wide structural variation system for dosage-based functional genomics offers a promising avenue for future research (Henry et al., 2015).

9.2 Potential for collaborative research projects and international cooperation

The poplar genome project has already demonstrated the benefits of international collaboration, as seen in the coordinated efforts to sequence and annotate the *Populus trichocarpa* genome (Tuskan et al., 2004). Future research can build on this foundation by fostering global partnerships to explore functional genomics and gene interactions in poplar. For instance, the PoplarGene network provides a valuable resource for mining functional information and can be leveraged for collaborative studies across different woody plant species (Liu et al., 2016). Additionally, initiatives like the Functional Annotation of All Salmonid Genomes (FAASG) can serve as a model for similar collaborative efforts in poplar genomics, promoting integrative research and resource sharing (Macqueen et al., 2017).

9.3 Funding and policy needs for sustained research in tree genomics

Sustained research in poplar genomics requires robust funding and supportive policies. The establishment of comprehensive functional genomics resources, such as cDNA libraries and microarray platforms, underscores the need for continued financial investment to maintain and expand these resources (Ralph et al., 2006). Furthermore, policies that encourage data sharing and open access to genomic databases will be crucial for advancing research. The creation of centralized databases, like the Poplar Functional Gene Database, can facilitate the organization and dissemination of experimental data, thereby enhancing the efficiency of research efforts (Si et al., 2015). To ensure long-term progress, funding agencies and policymakers must prioritize tree genomics and recognize its importance for both ecological and industrial applications.

10 Concluding Remarks

The analysis of the poplar genome has provided significant insights into the structural and functional aspects of this model woody plant. A comprehensive genome-wide analysis identified a substantial number of structural variations (SVs), including 7 889 deletions and 10 586 insertions, which affect 3 230 genes. These SVs are predominantly located in low-gene density regions and are associated with transposable elements, indicating their role in genome evolution and adaptation. Additionally, the creation of a functional gene network, PoplarGene, has facilitated the understanding of gene interactions and functional annotations, covering approximately 70% of the poplar genes. The development of full-length cDNA libraries and the identification of differentially expressed genes in response to insect feeding have further enriched the genomic resources available for poplar, aiding in the study of gene functions related to defense mechanisms.

This study has made substantial contributions to forestry science and genetic engineering by providing a detailed catalog of structural variations and their functional implications in poplar. The identification of genes involved in stress and pathogen resistance, as well as those associated with secondary cell wall biosynthesis, offers valuable targets for genetic improvement and breeding programs. The stability of transgene expression over multiple field seasons underscores the potential for reliable genetic modifications in poplar, which is crucial for both functional genomics studies and commercial forestry applications. Furthermore, the establishment of a dosage-based functional genomics system enables the exploration of gene dosage effects on quantitative traits and hybrid performance, paving the way for advanced breeding strategies.

Future genomic research should focus on expanding the functional annotation of the poplar genome, particularly in identifying and characterizing genes involved in key physiological processes such as dormancy, secondary growth, and long-term host-pest interactions. Integrating high-resolution RNA-Seq data with co-expression network analysis can provide deeper insights into the regulatory mechanisms governing these processes. Additionally, leveraging the PoplarGene network for cross-species functional annotation can enhance our understanding of gene functions in other woody plants, thereby broadening the scope of forestry research. Practical applications should prioritize the development of genetically engineered poplar varieties with enhanced resistance to biotic and abiotic stresses, improved wood quality, and optimized growth characteristics. This can be achieved through targeted manipulation of identified candidate genes and the application of advanced breeding techniques such as CRISPR/Cas9-mediated genome editing. By building on the foundational knowledge provided by this study, future research and practical applications can significantly advance the field of forestry science and contribute to sustainable forest management and wood production.

Acknowledgments

We appreciate the feedback from two anonymous peer reviewers on the manuscript of this study, whose careful evaluation and constructive suggestions have contributed to the improvement of the manuscript.

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.

Reference

Brunner A., Busov V., and Strauss S., 2004, Poplar genome sequence: functional genomics in an ecologically dominant plant species, Trends in Plant Science, 9(1): 49-56.
https://doi.org/10.1016/j.tplants.2003.11.006
PMid:14729219

Chen S., Yu Y., Wang X., Wang S., Zhang T., Zhou Y., He R., Meng N., Wang Y., Liu W., Liu Z., Liu J., Guo Q., Huang H., Sederoff R., Wang G., Qu G., and Chen S., 2023, Chromosome-level genome assembly of a triploid poplar *Populus alba* 'Berolinensis', Molecular Ecology Resources, 23: 1092-1107.
https://doi.org/10.1111/1755-0998.13770
PMid:36789493

Cho J., Jeon H., Kim M., Vo T., Kim J., Park E., Choi Y., Lee H., Han K., and Ko J., 2018, Wood forming tissue-specific bicistronic expression of *PdGA20ox1* and *PtrMYB221* improves both the quality and quantity of woody biomass production in a hybrid poplar, Plant Biotechnology Journal, 17(6): 1048-1057.
https://doi.org/10.1111/pbi.13036
PMid:30515982 PMCid:PMC6523601

Confalonieri M., Balestrazzi A., Bisoffi S., and Carbonera D., 2003, In vitro culture and genetic engineering of *Populus* spp.: synergy for forest tree improvement, Plant Cell, Tissue and Organ Culture, 72: 109-138.
https://doi.org/10.1023/A:1022265504775

Cruz F., Lagoa D., Mendes J., Rocha I., Ferreira E., Rocha M., and Dias Ó., 2019, *SamPler* - a novel method for selecting parameters for gene functional annotation routines, BMC Bioinformatics, 20: 454.
https://doi.org/10.1186/s12859-019-3038-4
PMid:31488049 PMCid:PMC6727554

Erwood S., Bily T., Lequyer J., Yan J., Gulati N., Brewer R., Zhou L., Pelletier L., Ivakine E., and Cohn R., 2021, Saturation variant interpretation using CRISPR prime editing, Nature Biotechnology, 40: 885-895.
https://doi.org/10.1038/s41587-021-01201-1
PMid:35190686

Fattel L., Psaroudakis D., Yanarella C., Chiteri K., Dostalik H., Joshi P., Starr D., Vu H., Wimalanathan K., and Lawrence-Dill C., 2021, Standardized genome-wide function prediction enables comparative functional genomics: a new application area for Gene Ontologies in plants, GigaScience, 11: giac023.
https://doi.org/10.1093/gigascience/giac023
PMid:35426911 PMCid:PMC9012101

Fitzpatrick M., and Keller S., 2015, Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation, Ecology Letters, 18(1): 1-16.
https://doi.org/10.1111/ele.12376
PMid:25270536

Ford K., McDonald D., and Mali P., 2019, Functional genomics via CRISPR-Cas, Journal of Molecular Biology, 431(1): 48-65.
https://doi.org/10.1016/j.jmb.2018.06.034
PMid:29959923 PMCid:PMC6309720

Geraldes A., Pang J., Thiessen N., Cezard T., Moore R., Zhao Y., Tam A., Wang S., Friedmann M., Birol I., Jones S., Cronk Q., and Douglas C., 2011, SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing, Molecular Ecology Resources, 11: 81-92.
https://doi.org/10.1111/j.1755-0998.2010.02960.x
PMid:21429165

Gerdes S., Yacoubi B., Bailly M., Blaby I., Blaby-Haas C., Jeanguenin L., Lara-Núñez A., Pribat A., Waller J., Wilke A., Overbeek R., Hanson A., and Crécy-Lagard V., 2011, Synergistic use of plant-prokaryote comparative genomics for functional annotations, BMC Genomics, 12: S2.
https://doi.org/10.1186/1471-2164-12-S1-S2
PMid:21810204 PMCid:PMC3223725

Hartenian E., and Doench J., 2015, Genetic screens and functional genomics using CRISPR/Cas9 technology, The FEBS Journal, 282(8): 1383-1393.
https://doi.org/10.1111/febs.13248
PMid:25728500

Henry I., Zinkgraf M., Groover A., and Comai L., 2015, A system for dosage-based functional genomics in poplar, Plant Cell, 27: 2370-2383.
https://doi.org/10.1105/tpc.15.00349
PMid:26320226 PMCid:PMC4815095

Hilton I., and Gersbach C., 2015, Enabling functional genomics with genome engineering, Genome Research, 25: 1442-1455.
https://doi.org/10.1101/gr.190124.115
PMid:26430154 PMCid:PMC4579329

Hsiao T., Revelles O., Chen L., Sauer U., and Vitkup D., 2009, Automatic policing of biochemical annotations using genomic correlations, Nature Chemical Biology, 6: 34-40.
https://doi.org/10.1038/nchembio.266
PMid:19935659 PMCid:PMC2935526

Huang D., Sherman B., and Lempicki R., 2008, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic Acids Research, 37: 1-13.
https://doi.org/10.1093/nar/gkn923
PMid:19033363 PMCid:PMC2615629

Joshi C., Bhandari S., Ranjan P., Kalluri U., Liang X., Fujino T., and Samuga A., 2004, Genomics of cellulose biosynthesis in poplars, The New Phytologist, 164(1): 53-61.
https://doi.org/10.1111/j.1469-8137.2004.01155.x
PMid:33873484

Kemmeren P., Berkum N., Vilo J., Bijma T., Donders R., Brazma A., and Holstege F., 2002, Protein interaction verification and functional annotation by integrated analysis of genome-scale data, Molecular Cell, 9(5): 1133-1143.
https://doi.org/10.1016/S1097-2765(02)00531-2
PMid:12049748

Kumar V., Hainaut M., Delhomme N., Mannapperuma C., Immerzeel P., Street N., Henrissat B., and Mellerowicz E., 2019, Poplar carbohydrate-active enzymes: whole-genome annotation and functional analyses based on RNA expression data, The Plant Journal, 99: 589-609.
https://doi.org/10.1111/tpj.14417
PMid:31111606 PMCid:PMC6852159

Lescot M., Rombauts S., Zhang J., Aubourg S., Mathé C., Jansson S., Rouzé P., and Boerjan W., 2004, Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements, Theoretical and Applied Genetics, 109: 10-22.
https://doi.org/10.1007/s00122-004-1621-0
PMid:15085260

Liu Q., Ding C., Chu Y., Chen J., Zhang W., Zhang B., Huang Q., and Su X., 2016, PoplarGene: poplar gene network and resource for mining functional information for genes from woody plants, Scientific Reports, 6: 31356.
https://doi.org/10.1038/srep31356
PMid:27515999 PMCid:PMC4981870

Liu S., Wang Z., Lan Y., He T., Xiong R., Wu C., Xiang Y., and Yan H., 2021, GEPSdb: the gene expression database of poplar under stress, The Plant Genome, 15(1): e20163.
https://doi.org/10.1002/tpg2.20163
PMid:34877793

Lu Q., Hu Y., Sun J., Cheng Y., Cheung K., and Zhao H., 2015, A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data, Scientific Reports, 5: 10576.
https://doi.org/10.1038/srep10576
PMid:26015273 PMCid:PMC4444969

Ma J., Wan D., Duan B., Bai X., Bai Q., Chen N., and Ma T., 2018, Genome sequence and genetic transformation of a widely distributed and cultivated poplar, Plant Biotechnology Journal, 17: 451-460.
https://doi.org/10.1111/pbi.12989
PMid:30044051 PMCid:PMC6335071

Macqueen D., Primmer C., Houston R., Nowak B., Bernatchez L., Bergseth S., Davidson W., Gallardo-Escárate C., Goldammer T., Guiguen Y., Iturra P., Kijas J., Koop B., Lien S., Maass A., Martin S., McGinnity P., Montecino M., Naish K., Nichols K., Ólafsson K., Omholt S., Palti Y., Plastow G., Rexroad C., Rise M., Ritchie R., Sandve S., Schulte P., Tello A., Vidal R., Vik J., Wargelius A., and Yáñez J., 2017, Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture, BMC Genomics, 18: 484.
https://doi.org/10.1186/s12864-017-3862-8
PMid:28655320 PMCid:PMC5488370

Manghwar H., Lindsey K., Zhang X., and Jin S., 2019, CRISPR/Cas system: recent advances and future prospects for genome editing, Trends in Plant Science, 24(12): 1102-1125.
https://doi.org/10.1016/j.tplants.2019.09.006
PMid:31727474

Nanjo T., Futamura N., Nishiguchi M., Igasaki T., Shinozaki K., and Shinohara K., 2004, Characterization of full-length enriched expressed sequence tags of stress-treated poplar leaves, Plant & Cell Physiology, 45(12): 1738-1748.
https://doi.org/10.1093/pcp/pci009
PMid:15653793

Nanjo T., Sakurai T., Totoki Y., Toyoda A., Nishiguchi M., Kado T., Igasaki T., Futamura N., Seki M., Sakaki Y., Shinozaki K., and Shinohara K., 2007, Functional annotation of 19 841 *Populus nigra* full-length enriched cDNA clones, BMC Genomics, 8: 448.
https://doi.org/10.1186/1471-2164-8-448
PMid:18053163 PMCid:PMC2222646

Park S., Oh S., and Han K., 2005, Large-scale computational analysis of poplar ESTs reveals the repertoire and unique features of expressed genes in the poplar genome, Molecular Breeding, 14: 429-440.
https://doi.org/10.1007/s11032-005-0603-5

Pinosio S., Giacomello S., Faivre-Rampant P., Taylor G., Jorge V., Paslier M., Zaina G., Bastien C., Cattonaro F., Marroni F., and Morgante M., 2016, Characterization of the poplar pan-genome by genome-wide identification of structural variation, Molecular Biology and Evolution, 33: 2706-2719.
https://doi.org/10.1093/molbev/msw161
PMid:27499133 PMCid:PMC5026262

Prasad T., Mohanty A., Kumar M., Sreenivasamurthy S., Dey G., Nirujogi R., Pinto S., Madugundu A., Patil A., Advani J., Manda S., Gupta M., Dwivedi S., Kelkar D., Hall B., Jiang X., Peery A., Rajagopalan P., Yelamanchi S., Solanki H., Raja R., Sathe G., Chavan S., Verma R., Patel K., Jain A., Syed N., Datta K., Khan A., Dammalli M., Jayaram S., Radhakrishnan A., Mitchell C., Na C., Kumar N., Sinnis P., Sharakhov I., Wang C., Gowda H., Tu Z., Kumar A., and Pandey A., 2017, Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes, Genome Research, 27: 133-144.
https://doi.org/10.1101/gr.201368.115
PMid:28003436 PMCid:PMC5204337

Proost S., Bel M., Vaneechoutte D., Peer Y., Inzé D., Mueller-Roeber B., and Vandepoele K., 2014, PLAZA 3.0: an access point for plant comparative genomics, Nucleic Acids Research, 43: D974-D981.
https://doi.org/10.1093/nar/gku986
PMid:25324309 PMCid:PMC4384038

Ralph S., Chun H., Cooper D., Kirkpatrick R., Kolosova N., Gunter L., Tuskan G., Douglas C., Holt R., Jones S., Marra M., and Bohlmann J., 2008, Analysis of 4 664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding, BMC Genomics, 9: 57.
https://doi.org/10.1186/1471-2164-9-57
PMid:18230180 PMCid:PMC2270264

Ralph S., Oddy C., Cooper D., Yueh H., Jancsik S., Kolosova N., Philippe R., Aeschliman D., White R., Huber D., Ritland C., Benoit F., Rigby T., Nantel A., Butterfield Y., Kirkpatrick R., Chun E., Liu J., Palmquist D., Wynhoven B., Stott J., Yang G., Barber S., Holt R., Siddiqui A., Jones S., Marra M., Ellis B., Douglas C., Ritland K., and Bohlmann J., 2006, Genomics of hybrid poplar (*Populus trichocarpa× deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar, Molecular Ecology, 15(5): 1275-1297.
https://doi.org/10.1111/j.1365-294X.2006.02824.x

Schulze S., and Lammers M., 2020, The development of genome editing tools as powerful techniques with versatile applications in biotechnology and medicine: CRISPR/Cas9, ZnF and TALE nucleases, RNA interference, and Cre/loxP, ChemTexts, 7(1): 3.
https://doi.org/10.1007/s40828-020-00126-7

Shalem O., Sanjana N., and Zhang F., 2015, High-throughput functional genomics using CRISPR–Cas9, Nature Reviews Genetics, 16: 299-311.
https://doi.org/10.1038/nrg3899
PMid:25854182 PMCid:PMC4503232

Shim D., Kim S., Choi Y., Song W., Park J., Youk E., Jeong S., Martinoia E., Noh E., and Lee Y., 2013, Transgenic poplar trees expressing yeast cadmium factor 1 exhibit the characteristics necessary for the phytoremediation of mine tailing soil, Chemosphere, 90(4): 1478-1486.
https://doi.org/10.1016/j.chemosphere.2012.09.044

Si J., Zhao X., Zhao X., and Wu R., 2015, Systematic functional genomics resource and annotation for poplar, IET Systems Biology, 9(4): 164-171.
https://doi.org/10.1049/iet-syb.2014.0047
PMid:26243833 PMCid:PMC8687384

Sterky F., Bhalerao R., Unneberg P., Segerman B., Nilsson P., Brunner A., Charbonnel-Campaa L., Lindvall J., Tandre K., Strauss S., Sundberg B., Gustafsson P., Uhlén M., Bhalerao R., Nilsson O., Sandberg G., Karlsson J., Lundeberg J., and Jansson S., 2004, A *Populus* EST resource for plant functional genomics, Proceedings of the National Academy of Sciences of the United States of America, 101(38): 13951-13956.
https://doi.org/10.1073/pnas.0401641101
PMid:15353603 PMCid:PMC518859

Tang W., Luo X., Nelson A., Collver H., and Kinken K., 2003, Functional genomics of wood quality and properties, Genomics, Proteomics & Bioinformatics, 1: 263-278.
https://doi.org/10.1016/S1672-0229(03)01032-5

Thakur A., Kumar P., Parmar N., Shandil R., Aggarwal G., Gaur A., and Srivastava D., 2021, Achievements and prospects of genetic engineering in poplar: a review, New Forests, 52: 889-920.
https://doi.org/10.1007/s11056-021-09836-3

Troyanskaya O., Dolinski K., Owen A., Altman R., and Botstein D., 2003, A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*), Proceedings of the National Academy of Sciences of the United States of America, 100: 8348-8353.
https://doi.org/10.1073/pnas.0832373100
PMid:12826619 PMCid:PMC166232

Tuskan G., DiFazio S., and Teichmann T., 2004, Poplar genomics is getting popular: the impact of the poplar genome project on tree research, Plant Biology, 6(1): 2-4.
https://doi.org/10.1055/s-2003-44715
PMid:15095128

Wang K., Li M., and Hakonarson H., 2010, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Research, 38: e164.
https://doi.org/10.1093/nar/gkq603
PMid:20601685 PMCid:PMC2938201

Yan L., Wei S., Wu Y., Hu R., Li H., Yang W., and Xie Q., 2015, High-efficiency genome editing in *Arabidopsis* using *YAO* promoter-driven CRISPR/Cas9 system, Molecular Plant, 8(12): 1820-1823.
https://doi.org/10.1016/j.molp.2015.10.004

Zhang R., Xu W., Shao S., and Wang Q., 2021, Gene silencing through CRISPR interference in bacteria: current advances and future prospects, Frontiers in Microbiology, 12: 635227.
https://doi.org/10.3389/fmicb.2021.635227
PMid:33868193 PMCid:PMC8044314