



OPEN ACCESS

EDITED BY

Lixin Wang,
Hebei Agricultural University, China

REVIEWED BY

Hui Ling,
Yulin Normal University, China

*CORRESPONDENCE

Yunpeng Cao
✉ xfcypeng@gxu.edu.cn
Guangcai Chen
✉ 335710534@qq.com

RECEIVED 14 October 2025

ACCEPTED 29 October 2025

PUBLISHED 17 November 2025

CITATION

Luo M, Wu H, Zhan D, Chen G and Cao Y
(2025) Rethinking *de novo* genes in
plants: mechanisms, methodological
progress, and future prospects.
Front. Plant Sci. 16:1724832.
doi: 10.3389/fpls.2025.1724832

COPYRIGHT

© 2025 Luo, Wu, Zhan, Chen and Cao. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Rethinking *de novo* genes in plants: mechanisms, methodological progress, and future prospects

Man Luo¹, Haibo Wu¹, Dinghua Zhan¹, Guangcai Chen^{2*}
and Yunpeng Cao^{3,4*}

¹School of Health and Nursing, Wuchang University of Technology, Wuhan, China, ²Guangxi Gaofeng State Owned Forest Farm, Nanning, China, ³Guangxi Colleges and Universities Key Laboratory for Cultivation and Utilization of Subtropical Forest Plantation, College of Forestry, Guangxi University, Nanning, China, ⁴Guangxi Key Laboratory of Forest Ecology and Conservation, College of Forestry, Guangxi University, Nanning, China

KEYWORDS

de novo, mechanisms, methodological, molecular features, multi-omics

Introduction

The emergence of novel genes represents a fundamental mechanism driving evolutionary innovation and adaptive evolution in living organisms (Xia et al., 2025). For decades, the prevailing paradigm in molecular evolution has held that new genes arise primarily through duplication and divergence of existing genes, horizontal gene transfer, or recombination events such as gene fusion and fission (Cao et al., 2024; Jiang et al., 2025; Xia et al., 2025). However, rapid advances in high-throughput sequencing and multi-species genomic data reveal that *de novo* genes (i.e. protein-coding genes arising from previously noncoding DNA) are far more common than once believed, fundamentally challenging the view that genetic novelty must originate solely from preexisting gene templates (Song et al., 2022; Cao et al., 2024; Xia et al., 2025). Initially considered evolutionary rarities or anomalies, *de novo* genes have now been identified across all domains of life, from bacteria to plants and animals (Broeils et al., 2023; Cao et al., 2024; Peng and Zhao, 2024; Xia et al., 2025). Plants, in particular, present an ideal system for studying *de novo* gene origination due to their expansive genomes, abundant non-coding regions, and high transposable element content, which collectively provide a rich substrate for the birth of novel genes (Xia et al., 2025). Recent large-scale comparative genomic studies have revealed that plant genomes harbor hundreds of lineage-specific genes lacking detectable homologs in closely related species, many of which show clear evidence of *de novo* origination from ancestral non-coding sequences (Song et al., 2022; Cao et al., 2024). The molecular signatures of plant *de novo* genes reveal intriguing patterns: they typically encode shorter proteins, lack recognizable conserved domains, and are enriched in intrinsically disordered regions (Song et al., 2022; Cao et al., 2024). While these features might appear suboptimal from a traditional protein evolution perspective, they may actually facilitate rapid functional exploration and adaptation to novel cellular contexts. Expression analyses consistently show that plant *de novo* genes exhibit highly restricted spatiotemporal

patterns, often being activated only during specific developmental stages, in particular tissues, or in response to environmental stresses—suggesting fine-tuned regulatory roles in adaptive responses (Song et al., 2022; Cao et al., 2024). Population genetic evidence increasingly supports the functional importance of *de novo* genes in plant adaptation (Cao et al., 2024; Zhao et al., 2024; Li et al., 2025). Several well-characterized examples demonstrate their contributions to key biological processes: the rice *OsDR10* gene confers pathogen resistance (Xiao et al., 2009), the Arabidopsis *AtQOS* gene regulates carbon-nitrogen metabolism and enhances disease resistance (Qi et al., 2019), Rosa *SCREP* regulates eugenol biosynthesis (Li et al., 2025), and numerous other *de novo* genes have been implicated in stress tolerance, reproductive success, and developmental regulation (Zhao et al., 2024). These discoveries underscore that *de novo* genes are not merely evolutionary noise but can provide substantive adaptive benefits. Despite major advances in *de novo* gene research, key challenges persist, including the need for high-quality genome assemblies, complex phylogenetic analyses, and multi-level functional validation for accurate identification, as well as the difficulty of distinguishing true *de novo* origins from rapid sequence divergence that obscures homology (Xia et al., 2025). Moreover, determining the functional significance of putative *de novo* genes and understanding how they integrate into existing gene regulatory networks represent ongoing scientific frontiers. This opinion article examines current understanding of *de novo* gene origination mechanisms in plants, evaluates methodological advances and limitations, and discusses implications for plant evolution and potential applications in crop improvement.

The current understanding and methodological advances in plant *de novo* gene studies

Mechanisms of origination: genome architecture and the role of transposable elements

Plant genomes provide an exceptionally fertile ground for *de novo* gene origination due to their unique architectural features. Large-scale comparative genomic analyses across diverse plant lineages reveal that extensive noncoding regions, comprising up to 85% of some plant genomes, harbor abundant cryptic open reading frames that can potentially evolve into functional genes (Zhao et al., 2024; Xia et al., 2025). This vast noncoding landscape, combined with frequent whole-genome duplications and chromosomal rearrangements characteristic of plant evolution, creates numerous opportunities for the emergence of novel coding sequences (Zhao et al., 2024; Xia et al., 2025). Transposable elements (TEs) play a particularly crucial role as catalysts for *de novo* gene birth in plants (Jin et al., 2021b; Zhao et al., 2024; Xia et al., 2025). Recent evidence demonstrates that TEs, which constitute 45–85% of many plant genomes, actively facilitate gene origination through multiple mechanisms (Jiang et al., 2022;

Pulido and Casacuberta, 2023; Cao et al., 2025). First, TE insertions can directly provide promoters, enhancers, and transcription factor binding sites that activate transcription of nearby noncoding sequences. Second, TEs mediate chromosomal rearrangements that bring together previously separated noncoding fragments, creating novel transcriptional units. Third, TE-induced epigenetic modifications can establish new chromatin states conducive to gene expression (Li et al., 2025; Xia et al., 2025). Analysis of rice, maize, and Arabidopsis genomes reveals that approximately 30–40% of recently originated *de novo* genes show clear associations with TE activity, either through direct sequence contribution or regulatory element donation (Xia et al., 2025). This TE-mediated mechanism appears particularly active during periods of environmental stress or genomic instability, potentially accelerating adaptive evolution through rapid gene innovation.

Molecular features: small, unstable proteins for rapid testing

Plant *de novo* genes exhibit distinctive molecular signatures that facilitate rapid functional exploration. These genes typically encode remarkably short proteins, often less than 100 amino acids, with high intrinsic disorder content and lacking recognizable conserved domains (Song et al., 2022; Cao et al., 2024; Xia et al., 2025). This structural “permissiveness” appears advantageous rather than detrimental—the abundance of disordered regions allows *de novo* proteins to escape strict folding constraints that govern canonical proteins, enabling them to act as flexible molecular probes capable of transient interactions and regulatory fine-tuning (Patiou et al., 2025; Xia et al., 2025). Studies in rice, Arabidopsis, and other plants consistently show that *de novo* proteins have lower intrinsic structural disorder (ISD) values, reduced GC content, and fewer secondary structure elements compared to conserved genes (Song et al., 2022; Cao et al., 2024; Peng and Zhao, 2024; Patiou et al., 2025). These properties enable rapid evolutionary testing of novel biochemical functions while minimizing the risk of misfolding and aggregation, essentially providing plants with a low-cost experimental platform for molecular innovation under selective pressures (Xia et al., 2025).

Expression and selective fate: evidence from population genomics and functional screens

Population genomic data suggest that plant *de novo* genes exhibit sharply restricted spatiotemporal expression, being chiefly induced during reproductive development or in response to environmental challenges like drought, pathogen exposure, and nutrient deficiency (Cao et al., 2024; Jiang et al., 2025; Xia et al., 2025). Large-scale transcriptomic surveys demonstrate that while most *de novo* genes display low expression levels compared to conserved genes, they show significant tissue specificity, with enrichment in reproductive tissues, suggesting roles as molecular fine-tuners of adaptive responses (Jin et al., 2021a; Song et al., 2022; Cao et al., 2024).

Selection-signature analyses (e.g., dN/dS ratios and population frequency distributions) show that *de novo* genes follow diverse evolutionary trajectories, with many genes (especially those involved in stress response and reproduction) being subject to positive or balancing selection (Kaessmann, 2010; Song et al., 2022; Cao et al., 2024; Xia et al., 2025). In addition, population studies also find that about 25%–30% of young genes become essential, such that their silencing is lethal (Li et al., 2025). However, many *de novo* genes are rapidly lost through genetic drift or negative selection, reflecting an ongoing evolutionary “trial-and-error” process (Van Oss and Carvunis, 2019; Xia et al., 2025). Functional validation through knockout experiments and CRISPR screens confirms that some *de novo* genes provide genuine adaptive advantages, such as the rice *OsDR10* conferring pathogen resistance and Arabidopsis *AtQQS* regulating metabolic networks (Xiao et al., 2009; Tanvir et al., 2022). Nevertheless, distinguishing truly functional *de novo* genes from transcriptional noise remains challenging, requiring convergent evidence from genomics, transcriptomics, proteomics, and experimental validation.

Methodological innovations: comparative genomics, multi-omics, and integrative frameworks

Recent methodological advances have revolutionized plant *de novo* gene identification and characterization. Progressive whole-genome alignment tools like Cactus now enable high-confidence synteny-based identification across divergent species, surpassing traditional BLAST-based approaches (Li et al., 2025; Xia et al., 2025). Multi-omics integration combining RNA-seq, Ribo-seq, proteomics, and metabolomics provides convergent evidence for gene functionality, addressing the challenge of distinguishing genuine *de novo* genes from transcriptional noise (Jin et al., 2021a; Song et al., 2022; Cao et al., 2024). Advanced computational frameworks incorporating deep learning (AlphaFold2) predict protein structures, revealing that some *de novo* proteins can achieve well-folded conformations despite lacking conserved domains (Li et al., 2025). Weighted gene co-expression network analysis (WGCNA) demonstrates how *de novo* genes integrate into existing regulatory networks (Jin et al., 2021a; Cao et al., 2024). Population genomics approaches using dN/dS ratios and selection signatures reveal adaptive evolution patterns (Song et al., 2022; Cao et al., 2024). These integrative pipelines, combining phylostratigraphy, expression profiling, and functional validation through CRISPR/Cas9, establish robust standards for *de novo* gene annotation and functional characterization in plants (Song et al., 2022; Li et al., 2025; Xia et al., 2025).

Challenges and hypotheses: strengths and weaknesses

Despite these advances, several unresolved problems demand attention. First, annotation errors and incomplete genome assemblies (especially widespread in polyploid and repetitive plant

genomes) affect the accuracy of gene age assignment and detection sensitivity (Xia et al., 2025). Second, phylostratigraphic approaches can overestimate *de novo* birth by failing to detect highly divergent homologs, while excessive stringency risks false negatives (Van Oss and Carvunis, 2019; Peng and Zhao, 2024). Third, not all detected ORFs possess biological function; some may reflect pervasive translation “noise,” and distinguishing functional *de novo* genes from translation byproducts remains technically challenging (Peng and Zhao, 2024; Xia et al., 2025). Misinterpretation is also a concern—identifying a recently fixed gene does not alone imply strong adaptive value, and function must still be established by knockout, phenotyping, or pathway analysis. The standards for *de novo* gene proof thus continue to shift toward convergence of evidence from genomics, transcriptomics, proteomics, and experimental approaches (Peng and Zhao, 2024).

The “proto-gene continuum” hypothesis, positing a spectrum from spurious ORFs to fully-fledged new genes, finds support in plant datasets: only a fraction of new sequences escape rapid loss, often after acquiring beneficial regulatory context through TE activity or environmental induction (Van Oss and Carvunis, 2019; Xia et al., 2025). Plant studies particularly demonstrate how the noncoding genome acts as a reservoir for rapid trait innovation, especially under strong selection or in adaptive radiations (Zhao et al., 2024; Xia et al., 2025). Importantly, recent research reveals the potential role of epigenetic state and regulatory plasticity in facilitating or constraining *de novo* gene emergence—topics that are rapidly gaining ground in the literature (Zhao et al., 2024; Li et al., 2025). Nevertheless, the field would benefit from more careful functional dissection, especially for genes found mostly in single accessions or populations. Scientific caution and explicit reporting of uncertainty are crucial to avoid over-attributing functions to recently emerged ORFs.

Discussion

Recent research has clarified the significant contribution of *de novo* genes to plant evolution and adaptation, although substantial challenges and open questions remain (Cao et al., 2024; Li et al., 2025). Plant genomes, rich in noncoding DNA and transposable elements, are particularly conducive to the emergence of new genes from previously noncoding regions (Tao et al., 2025). Transposable elements (TEs) create new genes through two primary mechanisms. First, they can generate genes from non-coding DNA by providing regulatory elements, like promoters, that activate adjacent sequences, or through the “exonization” of their own sequences. Second, they can modify existing genes. This occurs when a TE’s own gene is “molecularly domesticated” for a novel host function, as with the *RAG1* gene (Agrawal et al., 1998), or when TE insertion fuses host genes to create a new chimeric gene. Evidence shows that plant *de novo* genes often have highly specific expression and are rapidly induced by stress or distinct developmental processes (Zhao et al., 2024; Tóth et al., 2025; Xia et al., 2025), supporting the idea that they constitute a flexible, fast-evolving toolkit that helps plants handle novel environmental challenges. However, interpreting the functional impact of these genes requires caution. While experimental and population genetic

studies reveal that some *de novo* genes can provide adaptive advantages, the majority of candidate *de novo* genes remain uncharacterized or may even represent evolutionary transient entities. Distinguishing between genuinely functional *de novo* genes and those which are byproducts of pervasive translation remains an ongoing difficulty. Functional validation through knockout or overexpression studies is still available for only a minority of plant *de novo* genes (Xia et al., 2025).

Methodologically, the field continues to face significant obstacles, particularly in the annotation and validation of *de novo* gene candidates. High rates of sequence divergence, polyploidy, and limited genome annotation quality in many plant species can result in both false positives and false negatives when identifying and age-dating *de novo* genes (Xie et al., 2024; Xia et al., 2025). Recent advances such as deep co-linearity analysis, integrated multi-omics, and large-scale phylogenetic sampling all help to increase confidence, but robust community standards are still needed for declaring bona fide *de novo* genes (Xia et al., 2025). Despite these hurdles, ongoing research is moving towards a more nuanced view of genome innovation, where the *de novo* gene birth is not a rare accident, but a recurrent source of biological novelty. As functional genomics tools advance, systematic exploration of *de novo* gene roles in phenotypic traits, stress responses, and crop improvement will become increasingly feasible. Ultimately, integrating evolutionary, genomic, and ecological perspectives will be essential to fully understand the frequency, impact, and practical utility of *de novo* genes in plants. Future work should prioritize (1) standardized, multi-tier evidence frameworks reporting the status and confidence of candidate *de novo* genes, (2) integration of ecological, population, and molecular genetics, and (3) experimentally assessing the impact of *de novo* gene emergence on plant fitness and adaptation. In this way, plant *de novo* gene science can provide not only theoretical advances but also practical tools for sustainable agriculture and biological understanding.

Author contributions

ML: Writing – review & editing, Writing – original draft. HW: Writing – original draft, Writing – review & editing. DZ: Writing – original draft, Writing – review & editing. GC: Writing – review & editing, Writing – original draft. YC: Writing – review & editing, Writing – original draft.

References

- Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394, 744–751. doi: 10.1038/29457
- Broeils, L. A., Ruiz-Orera, J., Snel, B., Hubner, N., and Van Heesch, S. (2023). Evolution and implications of *de novo* genes in humans. *Nat. Ecol. Evol.* 7, 804–815. doi: 10.1038/s41559-023-02014-y
- Cao, Y., Feng, X., Ding, B., Huo, H., Abdullah, M., Hong, J., et al. (2025). Gap-free genome assemblies of two *Pyrus bretschneideri* cultivars and GWAS analyses identify

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was supported by a Guangxi “Bagui Young Talents” Special Fund.

Acknowledgments

Editors thank all the contributing authors in this Research Topic.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

a CCH zinc finger protein as a key regulator of stone cell formation in pear fruit. *Plant Commun.* 6, 101238. doi: 10.1016/j.xplc.2024.101238

Cao, Y., Hong, J., Zhao, Y., Li, X., Feng, X., Wang, H., et al. (2024). *De novo* gene integration into regulatory networks via interaction with conserved genes in peach. *Horticulture Res.* 11, uhac252. doi: 10.1093/hr/uhac252

Jiang, L., Li, X., Lyu, K., Wang, H., Li, Z., Qi, W., et al. (2025). Rosaceae phylogenomic studies provide insights into the evolution of new genes. *Hortic. Plant J.* 11, 389–405. doi: 10.1016/j.hpj.2024.02.002

- Jiang, L., Lin, M., Wang, H., Song, H., Zhang, L., Huang, Q., et al. (2022). Haplotype-resolved genome assembly of *Bletilla striata* (Thunb.) Reichb. f. to elucidate medicinal value. *Plant J.* 111, 1340–1353. doi: 10.1111/tpj.15892
- Jin, G., Ma, P.-F., Wu, X., Gu, L., Long, M., Zhang, C., et al. (2021a). New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos. *Mol. Biol. Evol.* 38, 5752–5768. doi: 10.1093/molbev/msab288
- Jin, G. H., Zhou, Y. L., Yang, H., Hu, Y. T., Shi, Y., Li, L., et al. (2021b). Genetic innovations: Transposable element recruitment and *de novo* formation lead to the birth of orphan genes in the rice genome. *J. Systematics Evol.* 59, 341–351. doi: 10.1111/jse.12548
- Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326. doi: 10.1101/gr.101386.109
- Li, Y., Li, R., Shang, J., Zhao, K., Sui, Y., Liu, Z., et al. (2025). A *de novo*-originated gene drives rose scent diversification. *Cell.* 20, 6121–6137. doi: 10.1016/j.cell.2025.08.011
- Patiou, C., Blassiau, C., Hatin, L., Eicholt, L. A., Corler, E., Ponitzki, C., et al. (2025). Pervasive translation of short open reading frames and *de novo* gene emergence in *Arabidopsis*. *bioRxiv*, 2025–2009.
- Peng, J., and Zhao, L. (2024). The origin and structural evolution of *de novo* genes in *Drosophila*. *Nat. Commun.* 15, 810. doi: 10.1038/s41467-024-45028-1
- Pulido, M., and Casacuberta, J. M. (2023). Transposable element evolution in plant genome ecosystems. *Curr. Opin. Plant Biol.* 75, 102418. doi: 10.1016/j.pbi.2023.102418
- Qi, M., Zheng, W., Zhao, X., Hohenstein, J. D., Kandel, Y., O'conner, S., et al. (2019). QQS orphan gene and its interactor NF-YC 4 reduce susceptibility to pathogens and pests. *Plant Biotechnol. J.* 17, 252–263. doi: 10.1111/pbi.12961
- Song, H., Guo, Z., Zhang, X., and Sui, J. (2022). *De novo* genes in *Arachis hypogaea* cv. Tifrunner: systematic identification, molecular evolution, and potential contributions to cultivated peanut. *Plant J.* 111, 1081–1095. doi: 10.1111/tpj.15875
- Tanvir, R., Ping, W., Sun, J., Cain, M., Li, X., and Li, L. (2022). AtQQS orphan gene and NtNF-YC4 boost protein accumulation and pest resistance in tobacco (*Nicotiana tabacum*). *Plant Sci.* 317, 111198. doi: 10.1016/j.plantsci.2022.111198
- Tao, X.-Y., Feng, S.-L., Yuan, L., Li, Y.-J., Li, X.-J., Guan, X.-Y., et al. (2025). Harnessing transposable elements for plant functional genomics and genome engineering. *Trends Plant Science.* 20, 1130–1146. doi: 10.1016/j.tplants.2025.03.007
- Tóth, D. M., Szeri, F., Ashaber, M., Muazu, M., Székvölgyi, L., and Arányi, T. (2025). Tissue-specific roles of *de novo* DNA methyltransferases. *Epigenet. Chromatin* 18, 5. doi: 10.1186/s13072-024-00566-2
- Van Oss, S. B., and Carvunis, A.-R. (2019). *De novo* gene birth. *PLoS Genet.* 15, e1008160. doi: 10.1371/journal.pgen.1008160
- Xia, S., Chen, J., Arsala, D., Emerson, J. J., and Long, M. (2025). Functional innovation through new genes as a general evolutionary process. *Nat. Genet.* 57, 295–309. doi: 10.1038/s41588-024-02059-0
- Xiao, W., Liu, H., Li, Y., Li, X., Xu, C., Long, M., et al. (2009). A rice gene of *de novo* origin negatively regulates pathogen-induced defense response. *PLoS One* 4, e4603. doi: 10.1371/journal.pone.0004603
- Xie, L., Gong, X., Yang, K., Huang, Y., Zhang, S., Shen, L., et al. (2024). Technology-enabled great leap in deciphering plant genomes. *Nat. Plants* 10, 551–566. doi: 10.1038/s41477-024-01655-6
- Zhao, L., Svetec, N., and Begun, D. J. (2024). *De novo* genes. *Annu. Rev. Genet.* 58, 211–232. doi: 10.1146/annurev-genet-111523-102413