Cell Genomics

CellPress
OPEN ACCESS

Perspective

Toward telomere-to-telomere genomics in Fabaceae: Unlocking comparative and functional insights into symbiotic nitrogen fixation

Yan Shi,[1,2] Chao Zhang,[1,2] Weicai Yang,[1,*] and Haifeng Wang[1,*]
[1]Yazhouwan National Laboratory, Sanya, Hainan 572025, China
[2]These authors contributed equally
*Correspondence: yangweicai@yzwlab.cn (W.Y.), wanghaifeng@yzwlab.cn (H.W.)
https://doi.org/10.1016/j.xgen.2026.101246

SUMMARY

The Fabaceae encompasses key agricultural species such as soybean, barrel medic (*Medicago*), and common bean, which are valued for their contributions to global food security and their ability to perform symbiotic nitrogen fixation. Although substantial progress has been made in sequencing economically important legumes, the taxonomic coverage remains uneven, with a disproportionate emphasis on model crops. Furthermore, most genome assemblies lack telomere-to-telomere (T2T) resolution, limiting insights into complex genomic regions and regulatory mechanisms. Emerging T2T technologies offer a transformative opportunity to overcome these limitations. By generating complete T2T assemblies for a representative range of Fabaceae species, including underrepresented lineages, researchers should obtain novel comparative and functional insights into the genetic and epigenetic bases of symbiotic nitrogen fixation. These high-resolution assemblies potentially facilitate the identification of conserved and divergent regulatory networks, shed light on the evolution of nodulation processes, and deepen our understanding of agronomically important traits.

INTRODUCTION

Symbiotic nitrogen fixation (SNF) is an interaction established between a host plant and nitrogen-fixing microorganisms. This interaction leads to the formation of a specialized organ in the host plant, the root nodule, which harbors intracellular bacteria.[1] Through this association, the microsymbiont converts atmospheric N_2 into NH_3, which is more readily taken up by the plant. In turn, the host plant reciprocally provides the microsymbiont C4 dicarboxylic acids, such as succinate and malate, as well as pyruvate, serving as carbon and energy sources. Plant species capable of SNF are restricted to a monophyletic lineage that is found in the orders Fabales (including the bean family), Fagales (including the birch family), Rosales (including the rose family), and Cucurbitales (including the Coriariaceae family).[2] This evolutionary branch is referred to as the nitrogen-fixing clade (NFC), which encompasses two forms: predominantly nodulation-based nitrogen fixation and, to a lesser extent, actinorhizal nitrogen fixation.[3–5]

The evolutionary history of SNF remains controversial. Studies show that novel complex traits have been generated through the reutilization of existing molecular mechanisms,[6] but the origin and the gains and losses of SNF remain uncertain.[7] This controversy stems primarily from the fact that there is a discontinuous phylogenetic distribution of nitrogen fixation across the NFC, which encompasses approximately 30,000 species. Additionally, evolutionary evidence remains elusive due to lineage loss, specifically (1) the extinction of ecologically or phylogenetically core species and (2) the disappearance of ancient lineages hypothesized to represent intermediate stages in the evolution of SNF.[2]

Currently, two primary evolutionary hypotheses have been proposed: a single-origin hypothesis (scenario I) and a multiple-origin model (scenario II).[8] Recent studies have provided evidence supporting the evolution of nitrogen fixation[9–12]; however, they have not fundamentally resolved which hypothesis is correct. The prevailing hypothesis of a single origin followed by extensive parallel losses has been challenged by a phylogenomic study proposing 16 independent origins and 10 subsequent losses.[9,10] Zhang et al. investigated the single origin of root nodules with genomic and transcriptomic data and proposed a gradual functionalization of nodule inception (*NIN*) from *NLP* precursors during evolution.[11] The common point of the two hypotheses is that multiple losses of the root nodule phenotype have occurred following its initial acquisition. Current evidence suggests that nodulation may have arisen from preexisting symbiotic programs, such as those underlying mycorrhizal associations, implying that few, if any, of its components evolved from novel genetic precursors. Traits associated with nodulation can be interpreted either as evidence of divergence from a common ancestor or as convergent evolution in distantly related lineages. Therefore, new approaches are needed to distinguish between a shared homologous origin and repeated independent assembly of similar genetic modules.

Considerable questions regarding the difference between these two hypotheses remain. The evolutionary formation of nodules integrates diverse biological processes found in non-nodulating

plants, including immune recognition, lateral root development, phytohormone signaling, and arbuscular mycorrhizal symbiosis. These functions establish a complex regulatory network that controls and coordinates nodule organogenesis and nitrogen fixation while facilitating the evolutionary emergence of novel genes and functions. However, hypotheses differ regarding the homology of root nodule symbiosis. The core issue is how these biological processes are reorganized and regulated in root nodules. This phenomenon is not caused by changes in one or a few genes but rather is the result of multiple evolutionary events.

The core transcription factor *NIN* may be key to unraveling this evolutionary puzzle.[6] The initial regulatory network for root nodule symbiosis, comprising 523 genes, originated from the common ancestor of nitrogen-fixing lineages. This network was constructed through a modular recruitment strategy, integrating and reorganizing genes from three major pathways: mycorrhizal symbiosis, nitrate response, and stress response.[12] Specifically, the ongoing debates center on three key issues: whether nodulation evolved once with multiple losses or independently multiple times; the extent to which pre-existing pathways were co-opted versus novel innovations emerged; the precise evolutionary timeline and selective pressures driving these multiple evolutionary events. The composition of the comparative genomic dataset used for analyses can significantly impact evolutionary studies on nodulation and nitrogen fixation. Therefore, more genetic data are needed to resolve these hypotheses and identify the core evolutionary mechanisms underlying the emergence and loss of nodulation. Clarifying the evolutionary relationships of SNF and identifying the minimal gene set required for its establishment have long been central research goals.

Fabaceae, the third-largest angiosperm family, comprising approximately 19,500 species across 765 genera, represents the primary NFC among plants. Therefore, its members serve as the primary resources for studying root nodulation-mediated nitrogen fixation.[13] Contemporary phylogenomic studies have resolved Fabaceae into six subfamilies: Cercidoideae, Detarioideae, Duparquetioideae, Dialioideae, Caesalpinioideae, and Papilionoideae (Figure 1). The most well-known nodulating model species, *Glycine max*, *Medicago truncatula*, and *Lotus japonicus*, belong to this family. Crops in the Fabaceae not only sustain global food security but also play critical roles in sustainable agriculture through nitrogen fixation.[14] This symbiotic process, mediated by root nodule formation with rhizobia, reduces reliance on synthetic nitrogen fertilizers, contributing to environmental sustainability and cost-effective crop production.[15]

Therefore, Fabaceae plants have a significant competitive advantage in nitrogen-deficient environments. The nodulating ability of Fabaceae corresponds to their phylogeny; nearly all Papilionoideae species exhibit nodulation, and the loss of this ability is more prevalent toward the basal nodes of the tree (Figure 1). These characteristics suggest that legumes occupy a unique and essential position in the analysis of the evolutionary process of SNF. Investigations of the diversity of nodulation and nitrogen fixation in Fabaceae plants may provide evidence of the evolutionary origin of root nodule-mediated nitrogen fixation. Additionally, integrating omics-level data analysis to elucidate the mechanisms underlying the loss of root nodule phenotypes could further inform this understanding.

THE TELOMERE-TO-TELOMERE GENOME AND ASSEMBLY STRATEGY

A telomere-to-telomere (T2T) genome assembly refers to a complete, gapless genome sequence that spans each chromosome from one telomere to the other without any gaps or missing regions.[17] High-quality reference genomes serve as a critical foundation for deciphering the genetic and regulatory mechanisms underlying morphological development through comparative genomic analyses. Currently, the construction of T2T genomes requires an integrated strategy that combines multiple long-read sequencing technologies with specialized assembly algorithms to resolve highly repetitive regions, including the centromeres and telomeres.[18] The foundational step involves generating high-accuracy, haplotype-resolved contigs; for example, using Pacific Biosciences (PacBio) high-fidelity (HiFi) sequencing, which produces reads of 15–25 kb with >99.9% single-read accuracy. To span the longest and most complex repetitive regions, ultra-long reads from Oxford Nanopore Technology (ONT), often exceeding 100 kb, are incorporated. These long reads are essential for bridging gaps and resolving structural variations in centromeric and pericentromeric regions. Subsequently, high-throughput chromosome conformation capture (Hi-C) sequencing is integrated to scaffold the assembled contigs into chromosome-scale sequences, providing long-range ordering and orientation information. The sequencing depth required for the three aforementioned sequencing technologies varies depending on genome size and genomic complexity; typical read coverage recommendations are approximately 60× for HiFi, 100× for ONT, and 100× for Hi-C.

The assembly workflow typically follows a multi-stage process.[19] First, high-fidelity PacBio HiFi reads and ONT ultra-long reads are assembled *de novo* using specialized tools such as Hifiasm.[20,21] Subsequently, Hi-C data are used to scaffold the assembly into chromosome-level sequences using Jucier, 3D-DNA, and HapHIC software (Figure 2).[22,23] The resulting contigs are then extended and merged using ONT ultra-long reads to close gaps, particularly within challenging heterochromatic regions, using tools such as TGS-GapCloser.[24] The final stage of T2T completion entails the precise identification and closure of residual gaps at telomeres and centromeres. This is achieved by mapping the canonical telomeric repeat motif (e.g., *Arabidopsis*-type "TTTAGGG") and analyzing centromere-specific satellite repeats and chromatin signatures using TeloComp, CentIER, and quarTeT tools.[25–27] Assembly completeness and accuracy are rigorously assessed using established metrics, including Benchmarking Universal Single-Copy Orthologs (BUSCO) and Quality Value (QV) scores, and validation of expected telomere and centromere counts per chromosome.[28,29] The inclusion of transcriptomes from roots and root nodules will further improve annotation accuracy using the BRAKER3 pipeline.[30] High-quality T2T reference genomes can be obtained using the workflow described above.

ADVANCES IN GENOMIC RESEARCH ON FABACEAE PLANTS

Genomic data represent the fundamental genetic code and provide the foundation for phenotypic diversity.[31] Genome

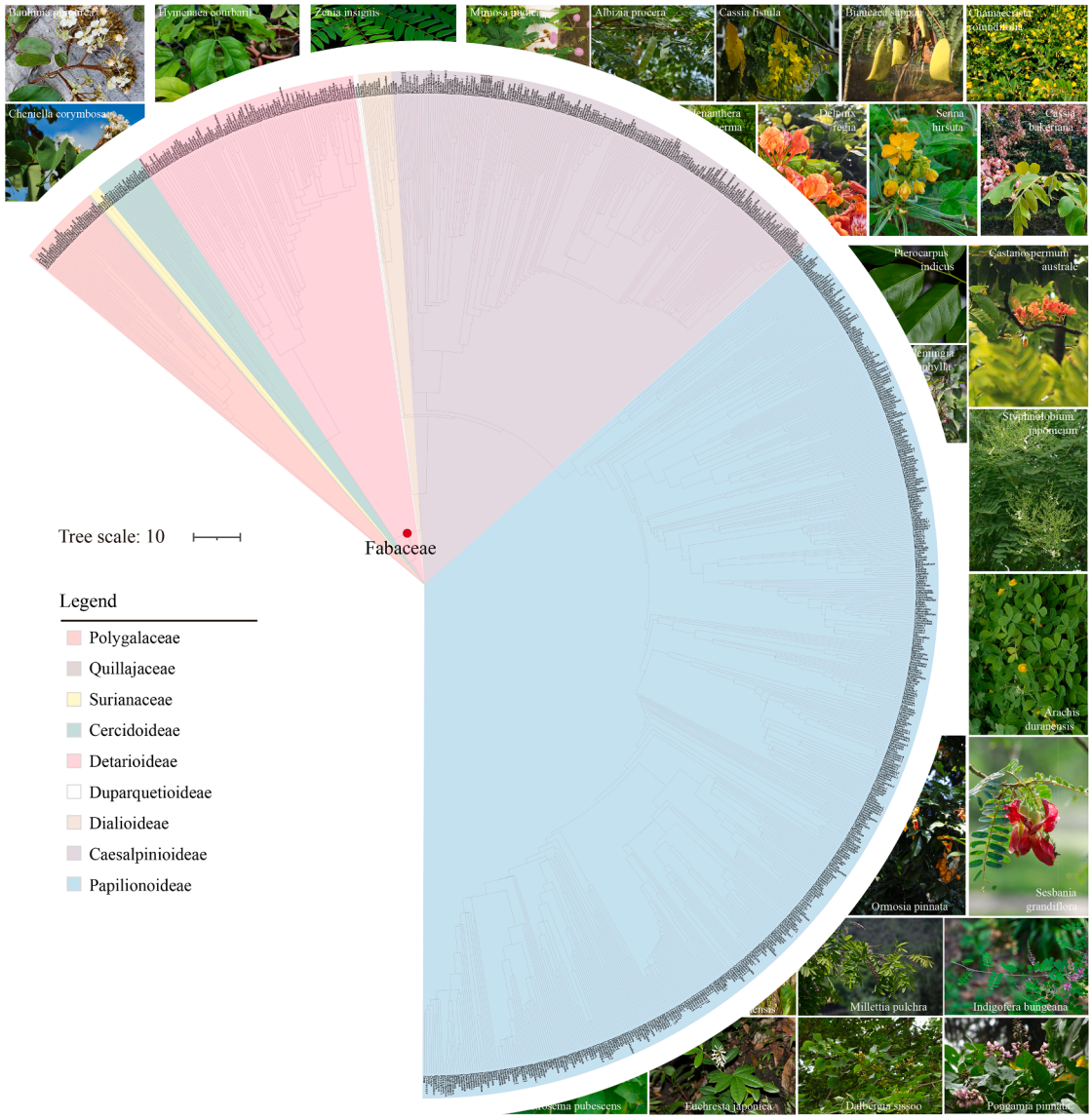

Figure 1. The diversity of leguminous plants

Leguminous plants exhibit remarkable diversity in their morphology, ecology, and genomic architecture, reflecting their widespread adaptation to diverse environmental conditions and their evolutionary success across a broad range of ecosystems. Information regarding the phylogenetic tree was obtained from http://www.timetree.org/home. The image was obtained from https://ppbc.iplant.cn/class with proper authorization.[16]

assemblies are commonly categorized by continuity and completeness into five hierarchical levels: contig, scaffold, chromosome, near-complete T2T (near-T2T), and T2T. The Fabaceae family has seen significant advances in the numbers of genomes sequenced over the past decade. To date, sequencing records are available for ~81 genera (~167 species). Among these, assemblies have been achieved at the contig level for ~15 species, at the scaffold level for ~40 species, at the chromosome level for >100 species, and at the T2T level for ~10 species.[32]

Genome sequencing of major Fabaceae crops has provided evidence to support the application of trait-regulating mechanisms for crop improvement. Legumes with high commercial value were the first to be sequenced: *G. max* (soybean), followed by *M. truncatula* (barrel medic), *Phaseolus vulgaris* (common bean), *Vigna angularis* (adzuki bean), *Pisum sativum* (pea), *Arachis hypogaea* (peanut), and *L. japonicus* (Japanese trefoil).[33–39] Most sequenced Fabaceae genomes have been assembled using a combination of short-read sequencing methods (e.g., Illumina) and long-read technologies (e.g., PacBio and ONT). The resulting assemblies reflect early-stage genome assembly quality. At this stage, while most genomic regions can be assembled into contigs, many remain unanchored to chromosomes, and numerous gap regions persist. Although these approaches have enabled chromosome-scale assemblies for some species, many genomes remain at

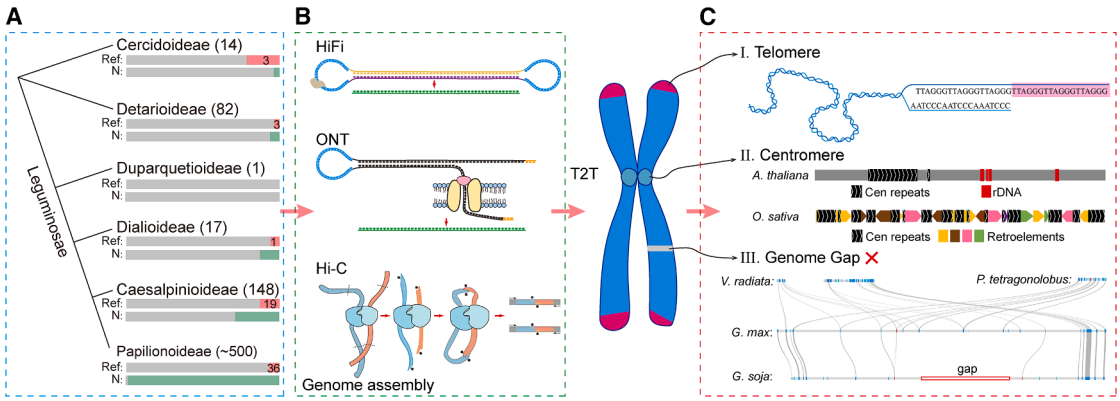

Figure 2. The process and advantages of T2T genome assembly

(A) Sequencing status and nodulation phenotypes of leguminous plants. Red bar chart, the number of sequenced genera; green bar chart, the approximate proportion of nitrogen-fixing species.

(B) T2T sequencing and assembly utilize 3 different methods to generate long reads: HiFi, Hi-C, and ONT.

(C) T2T sequencing can resolve telomere and centromere region functional analysis and identification of errors introduced by genome assembly gaps.

contig-level resolution, limiting their utility for comparative genomics.[32] The majority of genomic data are currently concentrated within the Papilionoideae subfamily of legumes, with the remaining five subfamilies lacking high-quality chromosome-scale genomes. Even contig-level reference genomes are absent in some subfamilies (Table 1). Consequently, the absence and uneven distribution of genomic data have hindered the full use of Fabaceae diversity in research on the evolution of SNF.[32,40]

Researchers are pursuing two main fronts: (1) improving the genome assembly quality and (2) expanding the taxonomic coverage of sequenced species. With respect to the first aspect, some studies are underway. For example, in *P. sativum,* an improved reference genome and a draft pan-genome, derived from second-generation sequencing data, have elucidated key genomic features of the species and shed light on the evolutionary dynamics underlying pea diversity.[55] Other chromosome-level genome assemblies and genetic variation maps of pea provide a valuable genomic resource for crop improvement. Additionally, the identification of trait-associated loci and candidate genes further supports this effort.[56] Due to errors in the pea genome assembly, genome-wide association study (GWAS) mapping has been inaccurate when comparing different versions of the reference genome.[37,55]

In *G. max*, the most commonly used reference genome, Wm82, has been updated to version 6, with continuous improvements in assembly quality to ensure accurate research in resource applications.[33,57,58] The high-quality reference genome assemblies of multiple soybean varieties have enabled the identification of single-nucleotide polymorphisms (SNP) and structural variation as well as analyses of population domestication analysis and phenotypic GWASs.[59–62] Similarly, an improved version of the genome of *M. truncatula* (Mt4.0), an important forage crop, has been released.[63] This advancement has propelled genome assembly into a new phase, ‘‘complete’’ genomes, which, compared to earlier assemblies of comparable quality, achieve near-complete chromosomal scaffolding and substantially reduce gap regions. However, complete genomes remain unable to resolve highly repetitive sequences and other structurally challenging regions.

Some studies are also underway with respect to the second aspect. Chromosome-level reference genomes for seven economically significant species within the Caesalpinioideae subfamily—*Mimosa bimucronata* (two-spurred mimosa), *Leucaena leucocephala* (white popinac), *Acacia confusa* (Taiwan acacia), *Delonix regia* (flame tree), *Albizia julibrissin* (silk tree), *Biancaea sappan* (sappan wood), and *Gleditsia sinensis* (zaojiao)—were used to analyze the evolutionary basis of chromosomal base number variation in this group.[49] Expanding the sequencing scope of legume genomes provides richer genomic resources for analyzing the diversity of root nodule nitrogen fixation across legumes, facilitating the identification of evolutionary changes in core gene families. Studies within these taxa have identified lineage-specific recent whole-genome duplication (WGD) events, characterized gene copies associated with secondary metabolites, and documented *de novo* gene origination events. They have also revealed genomic adaptations to extreme environments. Collectively, these findings may enhance our understanding of the genetic diversity of species in the Fabaceae.[49–52]

The studies mentioned above did not involve research on SNF. However, the genomic data generated can provide valuable support for future investigations. For example, the construction of a high-quality reference genome for *Zenia insignis*, combined with transcriptome analysis, revealed that the silencing of the nodulation factor receptor genes *NFR5*/*NFP* and the key transcription factor *NIN*, both essential for nodule initiation in legumes, may underlie the evolutionary loss of root nodules in this species.[53] The chromosomal-level genome sequencing and assembly of *Acacia melanoxylon* (black wattle) and comparative genomic analyses revealed dynamic expansion and contraction events in gene families associated with root nodule symbiosis and heartwood biosynthesis. Thus, these investigations refine the

Table 1. A summary of the highest-quality reference genomes for some key species across major Fabaceae plants

Subfamily	Species	Sequencing technologies	Assembly levels	Contig N50 (genome size)	BUSCO
Papilionoideae	*Glycine max*[41]	PacBio HiFi, Hi-C, ONT	T2T	48.76 Mb (0.93 Gb)	99.60%
Papilionoideae	*Medicago truncatula* A17[42]	PacBio HiFi, Hi-C, ONT	T2T	63.90 Mb (0.48 Gb)	99.19%
Papilionoideae	*Medicago truncatula* R108[42]	PacBio HiFi, Hi-C, ONT	T2T	54.40 Mb (0.41 Gb)	99.32%
Papilionoideae	*Sesbania cannabina*[43]	PacBio HiFi, Hi-C, ONT	T2T	180.56 Mb (2.04 Gb)	98.50%
Papilionoideae	*Arachis hypogaea*[44]	PacBio HiFi, Hi-C, ONT	T2T	126 Mb (2.6 Gb)	98.30%
Papilionoideae	*Phaseolus vulgaris*[45]	PacBio HiFi, Hi-C, ONT	T2T	55.11 Mb (0.55 Gb)	99.50%
Papilionoideae	*Vigna radiata*[46]	PacBio HiFi, Hi-C, ONT	T2T	46.00 Mb (0.47 Gb)	98.80%
Papilionoideae	*Lotus japonicus* MG-20[39]	PacBio, Hi-C	chromosome	2.40 Mb (0.51 Gb)	92.71%
Papilionoideae	*Pisum sativum*[37]	PacBio HiFi, Hi-C	chromosome	4.23 Mb (3.93 Gb)	99.20%
Papilionoideae	*Vigna umbellate*[47]	PacBio, Hi-C	chromosome	1.71 Mb (0.59 Gb)	97.20%
Caesalpinioideae	*Acacia melanoxylon*[48]	PacBio, Hi-C, ONT	chromosome	0.79 Mb (0.66 Gb)	90.00%
Caesalpinioideae	*Gleditsia sinensis*[49]	PacBio HiFi, Hi-C	chromosome	40.11 Mb (0.85 Gb)	98.90%
Caesalpinioideae	*Biancaea sappan*[49]	PacBio HiFi, Hi-C	chromosome	49.83 Mb (0.81 Gb)	99.30%
Caesalpinioideae	*Albizia julibrissin*[49]	PacBio HiFi, Hi-C	chromosome	41.11 Mb (0.66 Gb)	99.40%
Caesalpinioideae	*Delonix regia*[49]	PacBio HiFi, Hi-C	chromosome	34.15 Mb (0.54 Gb)	99.40%
Caesalpinioideae	*Acacia confusa*[49]	PacBio HiFi, Hi-C	chromosome	40.55 Mb (0.53 Gb)	99.20%
Caesalpinioideae	*Leucaena leucocephala*[49]	PacBio HiFi, Hi-C	chromosome	18.50 Mb (1.25 Gb)	99.70%
Caesalpinioideae	*Mimosa bimucronata*[49]	PacBio HiFi, Hi-C	chromosome	38.00 Mb (0.60 Gb)	99.40%
Cercidoideae	*Bauhinia variegata*[50]	PacBio, Hi-C	chromosome	4.55 Mb (0.32 Gb)	98.70%
Cercidoideae	*Phanera championii*[51]	PacBio HiFi, Hi-C	chromosome	27.71 Mb (0.38 Gb)	98.80%
Cercidoideae	*Cercis chinensis*[52]	PacBio HiFi, Hi-C	chromosome	33.27 Mb (0.35 Gb)	98.30%
Dialioideae	*Zenia insignis*[53]	PacBio HiFi, Hi-C	chromosome	10.03 Mb (0.39 Gb)	99.00%
Caesalpinioideae	*Eperua falcata*[54]	–	contig	–	–
Cercidoideae	*Cercis canadensis*[9]	–	contig	–	–

phylogenetic placement of *Acacia* within the Fabaceae family. Several gene families, including *RRS* and *PHYS*, have been identified as key regulators of both root nodule formation and broad ecological adaptability. These findings provides novel functional and evolutionary insights into the molecular mechanisms underlying SNF in legumes.[48]

Thus, genomic research in legumes has provided valuable insights into the gene families associated with nodulation, stress responses, and yield traits. Moreover, the core gene regulatory network of SNF is highly complex, and research on SNF, particularly quantitative and functional phenotypes, including root nodule presence/absence, nodule number, and nitrogenase activity, remains insufficient. Additionally, existing assemblies often lack the resolution to elucidate complex genomic regions, such as centromeres, telomeres, and other repetitive regions. These regions are difficult to assemble using data from the original chromosome-level assembly strategy. They may harbor loci associated with important agronomic traits and key biological pathways. Additionally, centromere dynamics and structural variations act synergistically to drive rapid morphological diversification in plants.[64] Thus, genomic investigations play a crucial role in elucidating evolutionary trajectories and functional adaptations. As the quality of genome assemblies significantly influences the accuracy of related research, T2T genomes can more effectively address assembly errors and reduce false positives in comparative genomics analyses.[65,66] However, most Fabaceae genomes lack T2T-level resolution, limiting their utility in comparative genomic and functional studies (Figure 2).[67]

THE ADVANTAGES OF T2T GENOMES IN FABACEAE PLANTS

Advances in sequencing and assembly technologies have driven the progressive refinement of genome assemblies from fragmented draft sequences toward T2T assemblies. Currently, T2T genome assembly is broadly categorized into two tiers: near-T2T and T2T assemblies. The key distinction lies in the proportion of chromosomes successfully assembled into T2T sequences relative to the full chromosomal complement. The first T2T genome assembly was achieved in human genomics, beginning with the complete resolution of individual chromosomes, starting with chromosome 8[68] and culminating in the assembly of a human T2T reference genome (T2T-CHM13).[69] In other animal species, a T2T genome assembly of a male *Sus scrofa* (domestic pig), designated T2T-pig1.0, has been released, leveraging the Chinese Wuzhishan miniature pig as a model. This assembly encompassing the full Y chromosome and all centromeric repeat arrays.[70] A T2T genome assembly of the *Capra hircus* (cashmere goat), designated T2T-goat1.0, has been released, with a total length of 2.86 Gb, encompassing a fully resolved Y chromosome of 20.96 Mb.[71]

Increasing numbers of T2T reference genomes have been generated for plant species. In *Arabidopsis thaliana*, researchers

reassembled the Col-0 genome to generate Col-CEN v1.2, a near-T2T reference genome. Chromosomes 1, 3, and 5 are fully T2T, while chromosomes 2 and 4 retain minor gaps predominantly located in pericentromeric and subtelomeric regions.[18] The Col-PEK assembly, a near-T2T reconstruction of the Col-0 ecotype, resolves the vast majority of gaps across the genome, including all five centromeric regions.[72] In *Oryza sativa* (rice), near-T2T reference genomes were first generated for two rice cultivars.[73] Subsequently, the first rice T2T reference genome, T2T-NIP (AGIS-1.0), was released, closing ~3% of the sequence gaps present in the prior reference IRGSP-1.0 and fully resolving the telomeric and centromeric regions across all 12 chromosomes. Similar T2T genomic advances have been demonstrated in *Zea mays* (maize), showcasing the transformative potential of T2T genomics in resolving long-standing questions in plant biology.[74]

Additionally, T2T reference genome assemblies have been completed for several economically important crops and fruits species: *Momordica charantia* (bitter melon)[75] and *Actinidia chinensis* (kiwifruit)[76] are near-T2T assemblies; whereas wild *Musa acuminata* (banana),[77] *Citrullus lanatus* (watermelon),[78] and wild *Fragaria vesca* (strawberry)[79] are T2T assemblies. These resources collectively provide a foundation for phenotypic mining and diversity identification. T2T reference genomes of polyploid, highly repetitive, and highly heterozygous species are being continuously generated, such as *Solanum tuberosum* (potato)[80] and *Triticum aestivum* (wheat),[81] demonstrating the feasibility of T2T assembly in such species.

The advances described above demonstrate that T2T assembly is now achievable for complex plant genomes, including those of the Fabaceae family.[82] Currently, only a handful of Fabaceae genomes, such as *G. max*, *M. truncatula*, and *Sesbania cannabina*,[41–43] have reached T2T standards, providing complete assemblies that resolve centromeres, telomeres, and large structural variations.

S. cannabina exhibits strong tolerance to alkaline soil and employs a distinct alkali tolerance pathway compared with both *Medicago sativa* (alfalfa) and *Sorghum bicolor* (sorghum). The T2T reference genome allowed the identification of four unique phosphate transporters (*PHT1* genes) induced by alkalinity in *S. cannabina*. This genetic distinction underlies *S. cannabina*'s unique alkali tolerance. However, the functional role of this gene remains experimentally unvalidated in other crop species, highlighting the potential for enhancing the adaptability of crops to phosphate deficiency in alkaline soils.[43] For the model accessions A17 and R108 of *M. truncatula*, high-quality T2T genome versions were obtained. Comparisons of the two genomes identified structural characteristics and evolutionary patterns of the centromere regions.[42]

Similarly, comparing a published soybean "complete" genome (ZH13v2)[83] and a T2T genome (ZH13-T2T) assembly was generated.[41] This genome sequence adds approximately 26 Mb of novel sequences and identified a previously unseen large inversion at the centromere, along with numerous other structural variations across the chromosomes.[41] Researchers constructed complete maps of centromeric satellite repeats from three T2T soybean genomes, elucidated their chromosome-specific evolution and interaction with CENH3 (centromere-specific histone H3), and proposed a model for centromeric DNA replication.[84] Plant centromeres typically consist of extensive repetitive DNA, but the functional centromere core is defined by the genomic region occupied by the centromere-specific histone CENH3. Only the T2T-grade assembly provides a complete, gapless representation of the entire centromere region, enabling rigorous investigation of centromere structure, evolution, and function. The T2T genome assembly of the wild soybean, *Glycine soja* YSD56, resistant to the soybean cyst nematode (one of the most destructive pathogens affecting soybean production), provides a high-quality genetic resource for studying disease resistance and soybean breeding.[85] Comparisons of the soybean ZN06 T2T genome and 341 germplasm resources resulted in the identification of three key genes: *GmFula*, *GmPdh1*, and *Dt1*, which are, respectively, associated with pod number, pod shattering, and growth habit.[86] These T2T assemblies have begun to provide critical insights into repetitive elements, stress-responsive pathways, and the evolution of nodulation in legumes.

Several studies have also focused on complex legume genomes, such as those that are polyploid and highly heterozygosis and have high repeat content and a large genome size. One study generated a T2T genome assembly of the tetraploid cultivated peanut (*A. hypogaea*) and investigated its evolutionary trajectory from its diploid wild progenitors (*Arachis duranensis* and *Arachis ipaensis*) to the cultivated species.[44] Some polyploid plants, particularly allopolyploids, have subgenomes, one of two evolutionarily distinct ancestral diploid genomes that collectively constitute the extant polyploid species. Accurate subgenome assignment is a foundational step in polyploid genomics and is dependent on robust identification of ancestral genome origins and reconstruction of their evolutionary relationships. In the case of peanut, most regions of subgenome A (subA) and subgenome B (subB) exhibit high collinearity, although multiple homologous exchange events were also detected.[44] Homologous recombination events reflect localized, intersubgenomic DNA exchanges that occurred during the evolutionary divergence of the two subgenomes. In *A. hypogaea*, such recombination is stochastic and can drive functional divergence among homoeologous gene pairs.

A closer examination of assembly metrics highlights disparities in genome quality across the Fabaceae. For instance, the T2T assemblies of *G. max* and *S. cannabina* have contig N50 values exceeding 20 Mb and BUSCO completeness scores above 98%.[41,43] By contrast, basal lineages such as *Cercis canadensis* (eastern redbud) and *Eperua falcata* (wallaba), are represented only by contig-level assemblies, with N50 values below 1 Mb and BUSCO scores under 90%.[9,54] These discrepancies underscore the need for targeted efforts to generate high-resolution assemblies of underrepresented lineages, which are essential for reconstructing the evolutionary history of Fabaceae and for identifying conserved and divergent regulatory networks. Efforts to generate additional T2T assemblies, particularly for basal and underrepresented lineages, are crucial to addressing these gaps.

Achieving complete, high-resolution assemblies for various Fabaceae species will enable comparative and functional insights into SNF and other key traits, supporting the development of innovative approaches for crop improvement and

sustainability. Such efforts should provide a robust foundation for comparative genomics and enable the identification of conserved pathways, novel regulatory elements, and lineage-specific adaptations that drive key traits. Through the genomic analysis of Fabaceae T2T genomes, we anticipate that the intrinsic mechanisms underlying the occurrence and loss of nodulation and nitrogen fixation phenotypes will also be clarified. This will enable a more accurate and comprehensive identification of the regulatory gene network for root nodules, the identification of core regulatory elements, and the detection of structural variations such as chromosomal rearrangements and inversions. Using these resources, the causes of nitrogen fixation formation likely will be comprehensively understood, and the origin model of root nodule nitrogen fixation (single or multiple times) can be clearly determined.

THE INITIATIVE OF FABACEAE T2T

Globally, initiatives such as the Earth BioGenome Project (EBP), Ruminant T2T (RT2T), and African Orphan Crops Consortium (AOCC) have underscored the importance of comprehensive genomic characterization across diverse plant and animal families.[87–89] These programs aim to generate high-quality assemblies to facilitate comparative studies and inform breeding programs. Similarly, we propose an initiative for the sequencing and assembly of Fabaceae plants, named the Fabaceae T2T (FT2T) initiative. The ultimate goals are to generate high-quality reference genomes for a broader range of legume species, addressing the current limitations of insufficient genome representation and uneven taxonomic distribution; to utilize these resources to elucidate the regulatory networks underlying SNF; as well as to explore potential functional roles of centromeres, telomeres, and other complex genomic regions in this process.

We have launched FT2T through a collaborative sampling and phenotyping network with leading Chinese institutions, including the Chinese Academy of Sciences, Chinese Academy of Tropical Agricultural Sciences, and China National Botanical Garden. In the current phase, our laboratory bears the core costs for high-fidelity genome sequencing and reference-grade assembly, while partner institutions provide well-documented germplasm accessions, rich associated metadata, and high-quality phenotypic data, including standardized assessments of nodulation and nitrogen fixation traits where biologically relevant. To ensure efficient coordination and prevent redundant effort, the FT2T currently operates through a tiered, mission-aligned consortium. It consists of a central sequencing, assembly, and quality control (SAQC) hub, hosted at the Yazhouwan National Laboratory in Sanya, China, that is responsible for generating, curating, and releasing reference-quality genomes and distributed partner teams focused on taxon-specific sample acquisition, trait characterization, and functional follow-up studies. At the current stage, species prioritization is coordinated by the corresponding authors together with core collaborating investigators, using explicit multi-dimensional criteria; as the initiative expands, this framework will transition into a broader steering structure with community participation.

Building on FT2T, high-quality T2T genome assemblies will be generated for approximately 100 phylogenetically diverse and carefully selected legume species. The primary consideration involves establishing clear species selection criteria that are tailored to the six subfamilies of Fabaceae, with distinct strategies applied to each. First, particular emphasis is placed on underrepresented basal clades, such as Cercidoideae and Detarioideae, with a representative species targeted for inclusion in each genus. Second, small subfamilies occupying basal phylogenetic positions, such as Duparquetioideae (comprising one genus) and Dialioideae (17 genera), are prioritized due to their significant evolutionary relevance—they represent early-diverging lineages that retain key ancestral traits and are crucial for understanding the origin, diversification, and character evolution of Fabaceae—despite their limited species diversity, and efforts are made to include all genera, each represented by one to two species. Third, Caesalpinioideae, the most taxonomically diverse subfamily, requires comprehensive sampling to include a broad representation of genera with careful inclusion of both nodulating and non-nodulating species, particularly when these traits co-occur within the same genus. Finally, within Papilionoideae, a representative set of core genera will be selected, with particular attention directed toward those that may have secondarily lost the capacity to form root nodules, such as *Castanospermum* and *Bowringia*.[90] During species selection, genome size, complexity, geographic origin, and phenotypic characteristics, including potential absences, must also be considered.

Diploid species with relatively small genome size (less than 3 Gb) will be prioritized to facilitate high-quality assembly and reduce associated costs.[91] When sufficient DNA for library construction cannot be obtained, optimized DNA extraction and sequencing library preparation methods will be applied. When phenotypic data are missing, field investigations will be conducted or seeds will be collected for cultivation, followed by phenotypic identification of the resulting seedlings. If all of these issues remain unresolved, the species will be excluded from the selection list. The selected species span multiple lineages within Fabaceae, including model species, crops, and wild relatives. Other representative taxa include *Callerya speciosa* (dragon's claw vine), *Trifolium pratense* (red clover), *Cicer arietinum* (chickpea), *Vigna unguiculata* (cowpea), *Astragalus membranaceus* (milk vetch), *Lupinus angustifolius* (narrow-leafed lupin), *Dalbergia odorifera* (rosewood), *Desmodium incanum* (beggarweed), *Psoralea corylifolia* (babchi), and *Erythrina variegata* (coral tree).

In the next phase, we anticipate that FT2T will become an international, contribution-based consortium. Because many Fabaceae lineages are geographically restricted to specific regions, building a truly representative family-wide T2T resource will require participation from researchers across multiple countries and regions. External groups will be able to participate by contacting the corresponding authors and contributing prioritized germplasm, voucher information, phenotype datasets, or downstream functional validation capacity. To ensure cross-species comparability, sequencing assembly and quality assessment will continue to be coordinated through the SAQC hub. To support transparency and reuse, raw sequencing data, genome assemblies, annotations, and associated metadata will be released through established public repositories with stable accession

numbers, version tracking, and clear quality metrics. In parallel, an FT2T database, which is currently under development, will provide an integrated entry point for assembly statistics, provenance records, phenotype descriptors, and comparative analyses across species. After release, datasets will be made freely available to the research community without consortium-restricted access. Authorship and consortium recognition will follow transparent, contribution-based principles that consider germplasm contribution, phenotyping, sequencing, assembly, annotation, and downstream biological validation.

At present, the core sequencing and assembly effort is supported by existing laboratory and project resources, whereas further expansion will depend on competitive national and international funding. Because FT2T is being developed in stages, not all further support has been secured in advance; rather, consortium growth will be matched to available resources and community participation.

Success will be rigorously evaluated against four primary metrics: (1) generation of reference-quality, near-T2T or T2T genome assemblies for a broad representation of Fabaceae species; (2) reconstruction of the gene-regulatory network underlying root nodule diversity; (3) identification and experimental validation of two to three evolutionarily conserved, functionally essential core genes underlying SNF; and (4) construction of a legume super-pan genome integrating genomic variation across the family. Through these efforts and attempts, a large number of high-quality T2T genomes can be generated, encompassing previously inaccessible genomic regions, such as centromeres, telomeres, and highly repetitive sequences, that are typically unresolved in conventional genome assemblies. These genomes provide robust data support for subsequent comparative genomic analyses integrating nodulation and nitrogen fixation phenotypes.

DISCUSSION

Integrating these newly generated genomes with previously published legume reference genomes enables systematic investigation of one of the most distinctive biological innovations in plant biology: SNF. Previous studies of the evolution of SNF have been hindered by the uneven quality of the genomes involved, which may introduce analytical errors.[9–11] These proposed high-resolution T2T genome assemblies will enable the elimination of analytical errors caused by genome assembly inaccuracies (Figure 2C). Gene gain and loss events associated with the emergence of the SNF phenotype, evolutionary trajectories of key genes, and the origin of novel genes can be systematically identified.[52,92] The minimal gene set required for root nodule SNF in legumes, as well as core regulatory genes beyond *NIN*, may be identified, leading to a more comprehensive understanding of the genetic and epigenetic foundations of SNF.

Furthermore, precise structural comparisons will be conducted for all identified SNF genes, along with analyses of their conserved non-coding regulatory elements. By integrating phenotypic data on root nodule nitrogen fixation with genome-wide structural variation profiles in leguminous plants, this integration may refine the mapping of structural variants associated with nitrogen fixation traits.[93] Additionally, the analysis of centromeric and telomeric regions in T2T assemblies elucidates genomic features that affect chromosome stability, recombination rates, and trait inheritance, improving our understanding of how genome structure shapes the evolution of legumes.[94] As WGD events and transposable elements are also significant drivers of genomic evolution,[90,95] the roles of centromeric regions, telomeric regions, and WGD events in the emergence of nitrogen fixation phenotypes remain to be elucidated.

The traditional "single-reference genome" misses much of a species' genetic variation. The pan-genome, built from multiple diverse individuals, captures more of its genetic repertoire and evolutionary history and helps identify core regulatory genes linked to phenotypes. Pan-genome studies have been conducted in various crops to investigate specific phenotypic traits. Research on the *Solanum* genus (which includes tomatoes and potatoes) has revealed that paralogous genes can serve as reservoirs for functional innovation in crop engineering.[96] Pan-genome analysis of 33 rice accessions uncovered previously undetected genomic variations.[97] In maize, a high-quality pan-genome has enabled the identification of structural variations such as *ZmASI3* that contribute to drought resistance.[98] Furthermore, integrating genomic and transcriptomic data from nine wheat varieties, a high-quality wheat pan-genome and pan-transcriptome have been constructed, providing insights into the core and variable genomic compartments, variety-specific genes, and transcriptional divergence.[99] T2T assemblies can enhance the integrity, accuracy, and comprehensiveness of pan-genome construction.

The combination of phased T2T assemblies and a super pan-genome is emerging as the ultimate reference framework for plant genomics.[100] For example, T2T assemblies of 27 *Citrullus* genotypes across seven species enabled construction of a high-quality pan-genome. It revealed strong structural variation (SV) associations with domestication traits—bitterness loss, sweetness, flesh color, and disease resistance—and weak associations with seed size, fruit shape, and maturity. These insights streamline marker-assisted breeding in watermelon.[101] Researchers assembled two T2T-level reference genomes from wild *Solanum aethiopicum* (African eggplant) and a cultivated accession (*Solanum melongena*). By integrating 22 representative germplasm accessions and four chromosome-level genomes, they constructed a representative pan-genome for the Asian eggplant and identified 463.94 Mb of non-reference sequences.[102] Wang et al. conducted a pan-genome analysis of edible legumes and explored several functions associated with nitrogen fixation.[103] These studies established a methodological framework for constructing T2T pan-genomes in legumes, leading to the identification of core, variable, and unique gene sets associated with root nodule formation and nitrogen fixation.

A T2T-level genome provides the highest-quality reference for integrating and comparing diverse omics datasets. As an example, comparative transcriptomics were used to examine symbiotic gene expression of nine host plants to investigate differences associated with various types of SNF.[104] Building on comparative genomics and transcriptomics, single-cell comparative transcriptomics addresses the challenge of high heterogeneity within the nodules. Studies have uncovered both universal and lineage-specific regulatory mechanisms and gene expression signatures in specific and rare cell populations with higher

sensitivity.[105] However, due to inherent limitations in library construction and read acquisition in single-cell transcriptome sequencing, single-cell analyses have more stringent requirements for reference genomes compared to bulk RNA sequencing (RNA-seq), and a T2T genome can provide the highest-quality reference to meet these demands.[106] Currently, single-cell studies of legume root nodules are limited to model species, and comparative transcriptomic analyses at the single-cell level across multiple legume species are lacking.[107–110] Simultaneously, we call for a legume plant database or symbiotic nitrogen-fixing plant database to be developed by integrating newly generated and published genomic data into a centralized resource.[32]

CONCLUSIONS

T2T genomics provides a transformative framework for understanding the complexities of the Fabaceae, reflecting both their ecological and agricultural significance. By generating T2T genome assemblies across various legume species, including underrepresented basal lineages, this study highlights the essential role of high-resolution genomic resources in addressing long-standing questions related to SNF and other agronomically important traits. The integration of T2T assemblies with comparative genomic and functional approaches, such as single-cell and spatial transcriptomics, offers opportunities to dissect conserved and divergent regulatory networks that govern nodulation and nitrogen fixation. These efforts will not only deepen our understanding of legume evolution but also enable targeted crop improvement strategies, supporting sustainable agriculture and global food security. By bridging taxonomic and genomic gaps, this study underscores the pivotal role of T2T genomics in establishing new benchmarks for exploring complex plant genomes. It provides a robust foundation for research integrating comparative, functional, and evolutionary genomics to harness the full potential of Fabaceae in addressing pressing global challenges.

ACKNOWLEDGMENTS

This work was supported by the Biological Breeding-National Science and Technology Major Project (2024ZD04079) and the National Key R&D Program of China (2025YFE0200400). The graphical abstract was created with BioRender (https://BioRender.com/1vckgk7).

AUTHOR CONTRIBUTIONS

H.W. and W.Y. conceived this project. H.W. supervised this project. Y.S. and C.Z. collected the materials and wrote the draft. H.W. and W.Y. revised this manuscript. All authors approved this manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors, particularly non-native English speakers, used ChatGPT and Copilot to refine sentences. After using this tool or service, the authors reviewed and edited the content extensively and take full responsibility for the content of the publication.

REFERENCES

1. Geurts, R., Xiao, T.T., and Reinhold-Hurek, B. (2016). What Does It Take to Evolve A Nitrogen-Fixing Endosymbiosis? Trends Plant Sci. *21*, 199–208. https://doi.org/10.1016/j.tplants.2016.01.012.
2. Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., and Martin, P.G. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA *92*, 2647–2651. https://doi.org/10.1073/pnas.92.7.2647.
3. Remigi, P., Zhu, J., Young, J.P.W., and Masson-Boivin, C. (2016). Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. Trends Microbiol. *24*, 63–75. https://doi.org/10.1016/j.tim.2015.10.007.
4. Ardley, J., and Sprent, J. (2021). Evolution and biogeography of actinorhizal plants and legumes: A comparison. J. Ecol. *109*, 1098–1121. https://doi.org/10.1111/1365-2745.13600.
5. Li, H.-L., Wang, W., Mortimer, P.E., Li, R.-Q., Li, D.-Z., Hyde, K.D., Xu, J.-C., Soltis, D.E., and Chen, Z.-D. (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. Sci. Rep. *5*, 14023. https://doi.org/10.1038/srep14023.
6. Doyle, J.J., Ren, J., Pawlowski, K., James, E.K., and Gao, Y. (2025). One versus many independent assemblies of symbiotic nitrogen fixation in flowering plants. Nat. Commun. *16*, 5345. https://doi.org/10.1038/s41467-025-60433-w.
7. Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., Lyons, E., Wang, M.-L., Chen, J., Biggers, E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. *47*, 1435–1442. https://doi.org/10.1038/ng.3435.
8. Battenberg, K., and Hayashi, M. (2022). Evolution of root nodule symbiosis: Focusing on the transcriptional regulation from the genomic point of view. Plant Biotechnol. *39*, 79–83. https://doi.org/10.5511/plantbiotechnology.22.0127a.
9. Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M.B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science *361*, eaat1743. https://doi.org/10.1126/science.aat1743.
10. Kates, H.R., O'Meara, B.C., LaFrance, R., Stull, G.W., James, E.K., Liu, S.-Y., Tian, Q., Yi, T.-S., Conde, D., Kirst, M., et al. (2024). Shifts in evolutionary lability underlie independent gains and losses of root-nodule symbiosis in a single clade of plants. Nat. Commun. *15*, 4262. https://doi.org/10.1038/s41467-024-48036-3.
11. Zhang, Y., Fu, Y., Xian, W., Li, X., Feng, Y., Bu, F., Shi, Y., Chen, S., van Velzen, R., Battenberg, K., et al. (2024). Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis. Plant Commun. *5*, 100671. https://doi.org/10.1016/j.xplc.2023.100671.
12. Liu, H., Hou, L., Lan, L., Zhang, R., Wang, D.-J., Feng, H., Chen, C.-Y., Ye, J.-J., Oyebanji, O.O., Chukwuma, E.C., et al. (2026). Evolution of root nodule symbiosis via paleopolyploidy and modular pathway rewiring. Cell Host Microbe *34*, 324–343.e13. https://doi.org/10.1016/j.chom.2026.01.001.
13. Ashley, N.E., and Mohammad, V. (2019). Advances in legume research in the genomics era. Aust. Syst. Bot. *32*, 459–483. https://doi.org/10.1071/SB19019.
14. Peoples, M.B., Brockwell, J., Herridge, D.F., Rochester, I.J., Alves, B.J.R., Urquiaga, S., Boddey, R.M., Dakora, F.D., Bhattarai, S., Maskey, S.L., et al. (2009). The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. Symbiosis *48*, 1–17. https://doi.org/10.1007/BF03179980.

15. Jensen, E.S., Peoples, M.B., Boddey, R.M., Gresshoff, P.M., Hauggaard-Nielsen, H., J R Alves, B., and Morrison, M.J. (2012). Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. Agron. Sustain. Dev. *32*, 329–364. https://doi.org/10.1007/s13593-011-0056-7.

16. Xie, G., Xuan, J., Liu, B., Luo, Y., Wang, Y., Zou, X., and Li, M. (2024). FlowerMate 2.0: Identifying plants in China with artificial intelligence. Innovation *5*, 100636. https://doi.org/10.1016/j.xinn.2024.100636.

17. Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., and Koren, S. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. *30*, 1291–1305. https://doi.org/10.1101/gr.263566.120.

18. Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., et al. (2021). The genetic and epigenetic landscape of the Arabidopsis centromeres. Science *374*, eabi7489. https://doi.org/10.1126/science.abi7489.

19. Aoyagi Blue, Y., Iimura, H., Sato, M.P., and Shirasawa, K. (2025). The impact of telomere-to-telomere genome assembly in the plant pan-genomics era. Breed. Sci. *75*, 3–12. https://doi.org/10.1270/jsbbs.24065.

20. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

21. Cheng, H., Asri, M., Lucas, J., Koren, S., and Li, H. (2024). Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. Nat. Methods *21*, 967–970. https://doi.org/10.1038/s41592-024-02269-8.

22. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95. https://doi.org/10.1126/science.aal3327.

23. Zeng, X., Yi, Z., Zhang, X., Du, Y., Li, Y., Zhou, Z., Chen, S., Zhao, H., Yang, S., Wang, Y., and Chen, G. (2024). Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat. Plants *10*, 1184–1200. https://doi.org/10.1038/s41477-024-01755-3.

24. Xu, M., Guo, L., Gu, S., Wang, O., Zhang, R., Peters, B.A., Fan, G., Liu, X., Xu, X., Deng, L., and Zhang, Y. (2020). TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience *9*, giaa094. https://doi.org/10.1093/gigascience/giaa094.

25. Huang, S.-B., Wu, J., Xu, Z.-J., Mo, W.-T., Yuan, S., Jiang, X.-Y., Wang, H.-F., and Xie, L. (2025). TeloComp: An efficient toolkit for accurate assembly of the telomeres in T2T genomes. Plant Commun. *6*, 101492. https://doi.org/10.1016/j.xplc.2025.101492.

26. Xu, D., Yang, J., Wen, H., Feng, W., Zhang, X., Hui, X., Yue, J., Xu, Y., Chen, F., and Pan, W. (2024). CentIER: Accurate centromere identification for plant genomes. Plant Commun. *5*, 101046. https://doi.org/10.1016/j.xplc.2024.101046.

27. Lin, Y., Ye, C., Li, X., Chen, Q., Wu, Y., Zhang, F., Pan, R., Zhang, S., Chen, S., Wang, X., et al. (2023). quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic. Res. *10*, uhad127. https://doi.org/10.1093/hr/uhad127.

28. Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E.M., and Kriventseva, E.V. (2025). OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res. *53*, D516–D522. https://doi.org/10.1093/nar/gkae987.

29. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. *21*, 245. https://doi.org/10.1186/s13059-020-02134-9.

30. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. *3*, lqaa108. https://doi.org/10.1093/nargab/lqaa108.

31. Kamble, M.A., Jha, S.K., and Sabale, P.M. (2025). Discovering the Genetic Code: Investigating Gene Expression Analysis and Genomic Sequencing. In Deep Learning and Computer Vision: Models and Biomedical Applications, *1*, U.N. Dulhare and E.H. Houssein, eds. (Springer Nature Singapore), pp. 47–61. https://doi.org/10.1007/978-981-96-1285-7_3.

32. Yu, Z., Lei, T., Yi, X., Hao, Y., Wu, S., Xiao, Z., Qu, J., Li, S., Wang, L., Li, Y., et al. (2025). LGRPv2: A high-value platform for the advancement of Fabaceae genomics. Plant Biotechnol. J. *23*, 4057–4075. https://doi.org/10.1111/pbi.70220.

33. Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature *463*, 178–183. https://doi.org/10.1038/nature08670.

34. Young, N.D., Debellé, F., Oldroyd, G.E.D., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F.X., Gouzy, J., Schoof, H., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature *480*, 520–524. https://doi.org/10.1038/nature10625.

35. Schmutz, J., McClean, P.E., Mamidi, S., Wu, G.A., Cannon, S.B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. Nat. Genet. *46*, 707–713.

36. Kang, Y.J., Satyawan, D., Shim, S., Lee, T., Lee, J., Hwang, W.J., Kim, S.K., Lestari, P., Laosatit, K., Kim, K.H., et al. (2015). Draft genome sequence of adzuki bean, *Vigna angularis*. Sci. Rep. *5*, 8069. https://doi.org/10.1038/srep08069.

37. Kreplak, J., Madoui, M.-A., Cápal, P., Novák, P., Labadie, K., Aubert, G., Bayer, P.E., Gali, K.K., Syme, R.A., Main, D., et al. (2019). A reference genome for pea provides insight into legume genome evolution. Nat. Genet. *51*, 1411–1422. https://doi.org/10.1038/s41588-019-0480-1.

38. Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M.K., Zhang, C., Chang, W.-C., Zhang, L., Zhang, X., Tang, R., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat. Genet. *51*, 865–876. https://doi.org/10.1038/s41588-019-0402-2.

39. Li, H., Jiang, F., Wu, P., Wang, K., and Cao, Y. (2020). A High-Quality Genome Sequence of Model Legume *Lotus japonicus* (MG-20) Provides Insights into the Evolution of Root Nodule Symbiosis. Genes *11*, 483. https://doi.org/10.3390/genes11050483.

40. Song, B., Ning, W., Wei, D., Jiang, M., Zhu, K., Wang, X., Edwards, D., Odeny, D.A., and Cheng, S. (2023). Plant genome resequencing and population genomics: Current status and future prospects. Mol. Plant *16*, 1252–1268. https://doi.org/10.1016/j.molp.2023.07.009.

41. Zhang, C., Xie, L., Yu, H., Wang, J., Chen, Q., and Wang, H. (2023). The T2T genome assembly of soybean cultivar ZH13 and its epigenetic landscapes. Mol. Plant *16*, 1715–1718. https://doi.org/10.1016/j.molp.2023.10.003.

42. Shen, L., Yi, C., Liu, Y., Han, F., and Feng, J. (2025). Two complete telomere-to-telomere Medicago genomes reveal the landscape and evolution of centromeres. Mol. Plant *18*, 1409–1412. https://doi.org/10.1016/j.molp.2025.07.016.

43. Luo, H., Wang, X., You, C., Wu, X., Pan, D., Lv, Z., Li, T., Zhang, D., Shen, Z., Zhang, X., et al. (2024). Telomere-to-telomere genome of the allotetraploid legume *Sesbania cannabina* reveals transposon-driven subgenome divergence and mechanisms of alkaline stress tolerance. Sci. China Life Sci. *67*, 149–160. https://doi.org/10.1007/s11427-023-2463-y.

44. Wang, X., Sun, Z., Qi, F., Zhou, Z., Du, P., Shi, L., Dong, W., Huang, B., Han, S., Pavan, S., et al. (2025). A telomere-to-telomere genome

assembly of the cultivated peanut. Mol. Plant *18*, 5–8. https://doi.org/10.1016/j.molp.2024.12.001.

45. Wang, Y., Hao, X., Chen, C., Wang, H., Gao, P., Yang, X., Dong, X., Qin, H., Li, M., Hou, S., et al. (2025). Telomere-to-telomere genome of common bean (*Phaseolus vulgaris* L., YP4). GigaScience *14*, giaf001. https://doi.org/10.1093/gigascience/giaf001.

46. Jia, K.-H., Li, G., Wang, L., Liu, M., Wang, Z.-W., Li, R.-Z., Li, L.-L., Xie, K., Yang, Y.-Y., Tian, R.-M., et al. (2025). Telomere-to-telomere, gap-free genome of mung bean (*Vigna radiata*) provides insights into domestication under structural variation. Hortic. Res. *12*, uhae337. https://doi.org/10.1093/hr/uhae337.

47. Francis, A., Singh, N.P., Singh, M., Sharma, P., Joshi, D.C., , et al.Kumar, D., Kumar, D., Basu, U., Bajaj, D., Varshney, N. (2023). The ricebean genome provides insight into Vigna genome evolution and facilitates genetic enhancement. Plant Biotechnol. J. *21*, 1522–1524. https://doi.org/10.1111/pbi.14075.

48. Zhang, R., Wang, Z., Zeng, B., Liu, J., Wang, H., Xie, J., and Fan, C. (2025). A chromosome-scale genome assembly of *Acacia melanoxylon* provides new insights into its root nodule and heartwood formation. Ind. Crops Prod. *226*, 120677. https://doi.org/10.1016/j.indcrop.2025.120677.

49. Chen, R., Meng, S., Wang, A., Jiang, F., Yuan, L., Lei, L., Wang, H., and Fan, W. (2024). The genomes of seven economic Caesalpinioideae trees provide insights into polyploidization history and secondary metabolite biosynthesis. Plant Commun. *5*, 100944. https://doi.org/10.1016/j.xplc.2024.100944.

50. Zhong, Y., Chen, Y., Zheng, D., Pang, J., Liu, Y., Luo, S., Meng, S., Qian, L., Wei, D., Dai, S., and Zhou, R. (2022). Chromosomal-level genome assembly of the orchid tree *Bauhinia variegata* (Leguminosae; Cercidoideae) supports the allotetraploid origin hypothesis of Bauhinia. DNA Res. *29*, dsac012. https://doi.org/10.1093/dnares/dsac012.

51. Lu, Y., Chen, X., Yu, H., Zhang, C., Xue, Y., Zhang, Q., and Wang, H. (2024). Haplotype-resolved genome assembly of *Phanera championii* reveals molecular mechanisms of flavonoid synthesis and adaptive evolution. Plant J. *118*, 488–505. https://doi.org/10.1111/tpj.16620.

52. Li, J., Shen, J., Wang, R., Chen, Y., Zhang, T., Wang, H., Guo, C., and Qi, J. (2023). The nearly complete assembly of the *Cercis chinensis* genome and Fabaceae phylogenomic studies provide insights into new gene evolution. Plant Commun. *4*, 100422. https://doi.org/10.1016/j.xplc.2022.100422.

53. Yu, H., Lu, Y., Zhang, C., Yang, W., Xie, H., Liu, H., and Wang, H. (2025). Genomic insights into the absence of root nodule formation and nitrogen fixation in *Zenia insignis*. J. Genet. Genomics *52*, 860–863. https://doi.org/10.1016/j.jgg.2024.12.001.

54. Brousseau, L., Fine, P.V.A., Dreyer, E., Vendramin, G.G., and Scotti, I. (2021). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata* Aubl. (Fabaceae). Mol. Ecol. *30*, 1136–1154. https://doi.org/10.1111/mec.15595.

55. Yang, T., Liu, R., Luo, Y., Hu, S., Wang, D., Wang, C., Pandey, M.K., Ge, S., Xu, Q., Li, N., et al. (2022). Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. Nat. Genet. *54*, 1553–1563. https://doi.org/10.1038/s41588-022-01172-2.

56. Liu, N., Lyu, X., Zhang, X., Zhang, G., Zhang, Z., Guan, X., Chen, X., Yang, X., Feng, Z., Gao, Q., et al. (2024). Reference genome sequence and population genomic analysis of peas provide insights into the genetic basis of Mendelian and other agronomic traits. Nat. Genet. *56*, 1964–1974. https://doi.org/10.1038/s41588-024-01867-8.

57. Valliyodan, B., Cannon, S.B., Bayer, P.E., Shu, S., Brown, A.V., Ren, L., Jenkins, J., Chung, C.Y.L., Chan, T.-F., Daum, C.G., et al. (2019). Construction and comparison of three reference-quality genome assemblies for soybean. Plant J. *100*, 1066–1082. https://doi.org/10.1111/tpj.14500.

58. Espina, M.J.C., Lovell, J.T., Jenkins, J., Shu, S., Sreedasyam, A., Jordan, B.D., Webber, J., Boston, L., Brůna, T., Talag, J., et al. (2024). Assembly, comparative analysis, and utilization of a single haplotype reference genome for soybean. Plant J. *120*, 1221–1235. https://doi.org/10.1111/tpj.17026.

59. Zhang, C., Shao, Z., Kong, Y., Du, H., Li, W., Yang, Z., Li, X., Ke, H., Sun, Z., Shao, J., et al. (2024). High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. Nat. Genet. *56*, 2247–2258. https://doi.org/10.1038/s41588-024-01901-9.

60. Yano, R., Li, F., Hiraga, S., Takeshima, R., Kobayashi, M., Toda, K., Umehara, Y., Kajiya-Kanegae, H., Iwata, H., Kaga, A., and Ishimoto, M. (2025). The genomic landscape of gene-level structural variations in Japanese and global soybean *Glycine max* cultivars. Nat. Genet. *57*, 973–985. https://doi.org/10.1038/s41588-025-02113-5.

61. Zhu, Z., Wang, Y., Liu, S., Wang, S., Li, J., Fang, C., Liu, Y., Yang, X., Tian, D., Song, S., and Tian, Z. (2025). Genomic atlas of 8,105 accessions reveals stepwise domestication, global dissemination, and improvement trajectories in soybean. Cell *188*, 6519–6535.e15. https://doi.org/10.1016/j.cell.2025.09.007.

62. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. *20*, 467–484. https://doi.org/10.1038/s41576-019-0127-1.

63. Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K.L., Yandell, M., Gundlach, H., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genom. *15*, 312. https://doi.org/10.1186/1471-2164-15-312.

64. Ma, W., Liu, Y., Wei, X., Zhang, X., Li, X., Liu, Z., Yuan, L., Li, G., Zhang, S., Yang, Q., et al. (2026). Gapless pangenome analyses reveal fast *Brassica rapa* subspeciation. Science *391*, eady7590. https://doi.org/10.1126/science.ady7590.

65. Ko, B.J., Lee, C., Kim, J., Rhie, A., Yoo, D.A., Howe, K., Wood, J., Cho, S., Brown, S., Formenti, G., et al. (2022). Widespread false gene gains caused by duplication errors in genome assemblies. Genome Biol. *23*, 205. https://doi.org/10.1186/s13059-022-02764-1.

66. Li, H., and Durbin, R. (2024). Genome assembly in the telomere-to-telomere era. Nat. Rev. Genet. *25*, 658–670. https://doi.org/10.1038/s41576-024-00718-w.

67. Garg, V., Bohra, A., Mascher, M., Spannagl, M., Xu, X., Bevan, M.W., Bennetzen, J.L., and Varshney, R.K. (2024). Unlocking plant genetics with telomere-to-telomere genome assemblies. Nat. Genet. *56*, 1788–1799. https://doi.org/10.1038/s41588-024-01830-7.

68. Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., et al. (2021). The structure, function and evolution of a complete human chromosome 8. Nature *593*, 101–107. https://doi.org/10.1038/s41586-021-03420-7.

69. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44–53. https://doi.org/10.1126/science.abj6987.

70. Luo, Y.-B., Huang, N., Zha, C.-W., Yang, L.-X., Xue, P.-X., Xu, Q., Yang, X.-Y., Zhang, L.-M., Wang, Y.-B., Chao, Z., et al. (2026). Telomere-to-telomere genome assembly of a male pig provides insight into population structure and selection for body stature. Nat. Genet. *58*, 195–205. https://doi.org/10.1038/s41588-025-02433-6.

71. Wu, H., Luo, L.-Y., Zhang, Y.-H., Zhang, C.-Y., Huang, J.-H., Mo, D.-X., Zhao, L.-M., Wang, Z.-X., Wang, Y.-C., He-Hua, E., et al. (2024). Telomere-to-telomere genome assembly of a male goat reveals variants associated with cashmere traits. Nat. Commun. *15*, 10041. https://doi.org/10.1038/s41467-024-54188-z.

72. Hou, X., Wang, D., Cheng, Z., Wang, Y., and Jiao, Y. (2022). A near-complete assembly of an *Arabidopsis thaliana* genome. Mol. Plant *15*, 1247–1250. https://doi.org/10.1016/j.molp.2022.05.014.

73. Song, J.-M., Xie, W.-Z., Wang, S., Guo, Y.-X., Koo, D.-H., Kudrna, D., Gong, C., Huang, Y., Feng, J.-W., Zhang, W., et al. (2021). Two gap-free reference genomes and a global view of the centromere architecture in rice. Mol. Plant *14*, 1757–1767. https://doi.org/10.1016/j.molp.2021.06.018.

74. Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., Li, T., Hu, J., Wang, K., Wang, C., Xin, B., et al. (2023). A complete telomere-to-telomere assembly of the maize genome. Nat. Genet. *55*, 1221–1231. https://doi.org/10.1038/s41588-023-01419-6.

75. Fu, A., Zheng, Y., Guo, J., Grierson, D., Zhao, X., Wen, C., Liu, Y., Li, J., Zhang, X., Yu, Y., et al. (2023). Telomere-to-telomere genome assembly of bitter melon (*Momordica charantia* L. var. abbreviata Ser.) reveals fruit development, composition and ripening genetic characteristics. Hortic. Res. *10*, uhac228. https://doi.org/10.1093/hr/uhac228.

76. Han, X., Zhang, Y., Zhang, Q., Ma, N., Liu, X., Tao, W., Lou, Z., Zhong, C., Deng, X.W., Li, D., and He, H. (2023). Two haplotype-resolved, gap-free genome assemblies for *Actinidia latifolia* and *Actinidia chinensis* shed light on the regulatory mechanisms of vitamin C and sucrose metabolism in kiwifruit. Mol. Plant *16*, 452–470. https://doi.org/10.1016/j.molp.2022.12.022.

77. Belser, C., Baurens, F.-C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., Hřibová, E., Doležel, J., et al. (2021). Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. Commun. Biol. *4*, 1047. https://doi.org/10.1038/s42003-021-02559-3.

78. Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., Xu, B., Tian, Y., Sun, Y., Li, B., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. Mol. Plant *15*, 1268–1284. https://doi.org/10.1016/j.molp.2022.06.010.

79. Zhou, Y., Xiong, J., Shu, Z., Dong, C., Gu, T., Sun, P., He, S., Jiang, M., Xia, Z., Xue, J., et al. (2023). The telomere-to-telomere genome of *Fragaria vesca* reveals the genomic evolution of Fragaria and the origin of cultivated octoploid strawberry. Hortic. Res. *10*, uhad027. https://doi.org/10.1093/hr/uhad027.

80. Yang, X., Zhang, L., Guo, X., Xu, J., Zhang, K., Yang, Y., Yang, Y., Jian, Y., Dong, D., Huang, S., et al. (2023). The gap-free potato genome assembly reveals large tandem gene clusters of agronomical importance in highly repeated genomic regions. Mol. Plant *16*, 314–317. https://doi.org/10.1016/j.molp.2022.12.010.

81. Liu, S., Li, K., Dai, X., Qin, G., Lu, D., Gao, Z., Li, X., Song, B., Bian, J., Ren, D., et al. (2025). A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. Nat. Genet. *57*, 1008–1020.

82. Cheung, T.-Y., Kafoutchoni, K.M., Agoyi, E.E., Chan, T.-F., and Chapman, M.A. (2025). Exceptionally low genomic diversity in the underutilised legume Kersting's groundnut. Nat. Commun. *16*, 5183. https://doi.org/10.1038/s41467-025-60494-x.

83. Shen, Y., Du, H., Liu, Y., Ni, L., Wang, Z., Liang, C., and Tian, Z. (2019). Update soybean Zhonghuang 13 genome to a golden reference. Sci. China Life Sci. *62*, 1257–1260. https://doi.org/10.1007/s11427-019-9822-2.

84. Huang, Y., Guan, E., Song, S., Koo, D.-H., Schmidt, M.A., Su, H., Chen, C., and Zhang, J. (2026). Genetic diversity and architectural dynamics of soybean centromeres. Genome Biol. *27*, 17. https://doi.org/10.1186/s13059-025-03924-9.

85. Lian, Y., Wu, Y., Li, C., Wei, H., Du, P., Li, J., Lei, C., Li, H., Wang, S., Zhang, H., et al. (2025). A telomere-to-telomere genome of wild soybean with resistance to soybean cyst nematode X12. Sci. Data *12*, 1412. https://doi.org/10.1038/s41597-025-05741-y.

86. Liu, N., Feng, W., Zhang, G., Gao, P., Lian, L., Yuan, J., Liu, X., Ni, X., Wang, H., Ou, J., et al. (2025). Genomic and population evidence uncovers divergent improvement of vegetable soybean from grain soybean. Mol. Plant *18*, 1094–1097. https://doi.org/10.1016/j.molp.2025.06.009.

87. Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., et al. (2018). Earth BioGenome Project: Sequencing life for the future of life. Proc. Natl. Acad. Sci. USA *115*, 4325–4333. https://doi.org/10.1073/pnas.1720115115.

88. Kalbfleisch, T.S., McKay, S.D., Murdoch, B.M., Adelson, D.L., Almansa-Villa, D., Becker, G., Beckett, L.M., Benítez-Galeano, M.J., Biase, F., Casey, T., et al. (2024). The Ruminant Telomere-to-Telomere (RT2T) Consortium. Nat. Genet. *56*, 1566–1573. https://doi.org/10.1038/s41588-024-01835-2.

89. Jamnadass, R., Mumm, R.H., Hale, I., Hendre, P., Muchugi, A., Dawson, I.K., Powell, W., Graudal, L., Yana-Shapiro, H., Simons, A.J., and Van Deynze, A. (2020). Enhancing African orphan crops with genomics. Nat. Genet. *52*, 356–360. https://doi.org/10.1038/s41588-020-0601-x.

90. Zhao, Y., Zhang, R., Jiang, K.-W., Qi, J., Hu, Y., Guo, J., Zhu, R., Zhang, T., Egan, A.N., Yi, T.-S., et al. (2021). Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. Mol. Plant *14*, 748–773. https://doi.org/10.1016/j.molp.2021.02.006.

91. Li, J., Liu, Z., You, C., Qi, Z., You, J., Grover, C.E., Long, Y., Huang, X., Lu, S., Wang, Y., et al. (2024). Convergence and divergence of diploid and tetraploid cotton genomes. Nat. Genet. *56*, 2562–2573. https://doi.org/10.1038/s41588-024-01964-8.

92. Zhang, L., Liu, Y., Huang, Y., Zhang, Y., Fu, Y., Xiao, Y., Chen, S., Zhang, K., and Cheng, F. (2025). Solanaceae pan-genomes reveal extensive fractionation and functional innovation of duplicated genes. Plant Commun. *6*, 101231. https://doi.org/10.1016/j.xplc.2024.101231.

93. Zhang, A., Kong, T., Sun, B., Qiu, S., Guo, J., Ruan, S., Guo, Y., Guo, J., Zhang, Z., Liu, Y., et al. (2024). A telomere-to-telomere genome assembly of Zhonghuang 13, a widely-grown soybean variety from the original center of *Glycine max*. Crop J. *12*, 142–153. https://doi.org/10.1016/j.cj.2023.10.003.

94. Herklotz, V., Zhang, M., Nascimento, T., Kalfusová, R., Lunerová, J., Fuchs, J., Harpke, D., Huettel, B., Pfordt, U., Wissemann, V., et al. (2025). Bimodal centromeres in pentaploid dogroses shed light on their unique meiosis. Nature *643*, 148–157. https://doi.org/10.1038/s41586-025-09171-z.

95. Tao, X.-Y., Feng, S.-L., Yuan, L., Li, Y.-J., Li, X.-J., Guan, X.-Y., Chen, Z.-H., and Xu, S.-C. (2025). Harnessing transposable elements for plant functional genomics and genome engineering. Trends Plant Sci. *30*, 1130–1146. https://doi.org/10.1016/j.tplants.2025.03.007.

96. Benoit, M., Jenike, K.M., Satterlee, J.W., Ramakrishnan, S., Gentile, I., Hendelman, A., Passalacqua, M.J., Suresh, H., Shohat, H., Robitaille, G.M., et al. (2025). Solanum pan-genetics reveals paralogues as contingencies in crop engineering. Nature *640*, 135–145. https://doi.org/10.1038/s41586-025-08619-6.

97. Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell *184*, 3542–3558.e16. https://doi.org/10.1016/j.cell.2021.04.046.

98. Yang, S., Wang, Y., Huang, Q., Wang, M., Wang, S., Fu, X., Zhu, C., Cheng, J., Liu, S., Yang, Z., et al. (2025). A pangenome of maize provides genetic insights into drought resistance. Nat. Genet. *57*, 2831–2841. https://doi.org/10.1038/s41588-025-02378-w.

99. White, B., Lux, T., Rusholme-Pilcher, R., Juhász, A., Kaithakottil, G., Duncan, S., Simmonds, J., Rees, H., Wright, J., Colmer, J., et al. (2025). De novo annotation reveals transcriptomic complexity across the hexaploid wheat pan-genome. Nat. Commun. *16*, 8538. https://doi.org/10.1038/s41467-025-64046-1.

100. Zhang, J., Zheng, Y., and Chen, F. (2026). Phased telomere-to-telomere super-pangenome: definitive reference genome in plants. Trends Plant Sci. *31*, 266–269. https://doi.org/10.1016/j.tplants.2025.11.002.

101. Zhang, Y., Zhao, M., Tan, J., Huang, M., Chu, X., Li, Y., Han, X., Fang, T., Tian, Y., Jarret, R., et al. (2024). Telomere-to-telomere Citrullus

super-pangenome provides direction for watermelon breeding. Nat. Genet. *56*, 1750–1761. https://doi.org/10.1038/s41588-024-01823-6.

102. Yu, C., Li, W., Jiang, Y., Yang, Q., Gan, G., Cai, I., Li, W., and Wang, Y. (2026). Graph pangenome advances genetic discoveries and the improvement of eggplant. Hortic. Res. *13*, uhaf248. https://doi.org/10.1093/hr/uhaf248.

103. Wang, L., Jiang, X., Jiao, W., Mao, J., Ye, W., Cao, Y., Chen, Q., and Song, Q. (2025). Pangenome analysis provides insights into legume evolution and breeding. Nat. Genet. *57*, 2052–2061. https://doi.org/10.1038/s41588-025-02280-5.

104. Libourel, C., Keller, J., Brichet, L., Cazalé, A.-C., Carrère, S., Vernié, T., Couzigou, J.-M., Callot, C., Dufau, I., Cauet, S., et al. (2023). Comparative phylotranscriptomics reveals ancestral and derived root nodule symbiosis programmes. Nat. Plants *9*, 1067–1080. https://doi.org/10.1038/s41477-023-01441-w.

105. Cervantes-Pérez, S.A., Zogli, P., Amini, S., Thibivilliers, S., Tennant, S., Hossain, M.S., Xu, H., Meyer, I., Nooka, A., Ma, P., et al. (2024). Single-cell transcriptome atlases of soybean root and mature nodule reveal new regulatory programs that control the nodulation process. Plant Commun. *5*, 100984. https://doi.org/10.1016/j.xplc.2024.100984.

106. Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S., and Theis, F.J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nat. Commun. *10*, 390. https://doi.org/10.1038/s41467-018-07931-2.

107. Liu, Z., Kong, X., Long, Y., Liu, S., Zhang, H., Jia, J., Cui, W., Zhang, Z., Song, X., Qiu, L., et al. (2023). Integrated single-nucleus and spatial transcriptomics captures transitional states in soybean nodule maturation. Nat. Plants *9*, 515–524. https://doi.org/10.1038/s41477-023-01387-z.

108. Liu, Z., Yang, J., Long, Y., Zhang, C., Wang, D., Zhang, X., Dong, W., Zhao, L., Liu, C., Zhai, J., and Wang, E. (2023). Single-nucleus transcriptomes reveal spatiotemporal symbiotic perception and early response in Medicago. Nat. Plants *9*, 1734–1748. https://doi.org/10.1038/s41477-023-01524-8.

109. Cervantes-Pérez, S.A., Thibivilliers, S., Laffont, C., Farmer, A.D., Frugier, F., and Libault, M. (2022). Cell-specific pathways recruited for symbiotic nodulation in the *Medicago truncatula* legume. Mol. Plant *15*, 1868–1888. https://doi.org/10.1016/j.molp.2022.10.021.

110. Ye, Q., Zhu, F., Sun, F., Wang, T.-C., Wu, J., Liu, P., Shen, C., Dong, J., and Wang, T. (2022). Differentiation trajectories and biofunctions of symbiotic and un-symbiotic fate cells in root nodules of *Medicago truncatula*. Mol. Plant *15*, 1852–1867. https://doi.org/10.1016/j.molp.2022.10.019.