*Nucleic Acids Research*, 2026, gkag423
https://doi.org/10.1093/nar/gkag423
**Web Server issue**

OXFORD

# T2T-Hub: a central platform for analyzing plant and animal telomere-to-telomere genomes

**Haoyu Chao [1,†], Zhonghao Ruan [2,†], Zhuojin Li [2,†], Xinkai Zhou [2], Shilong Zhang [1], Xingxing Shen [3], Cong Feng [1,4,*], Dijun Chen [2,*], Ming Chen [1,4,5,*]**

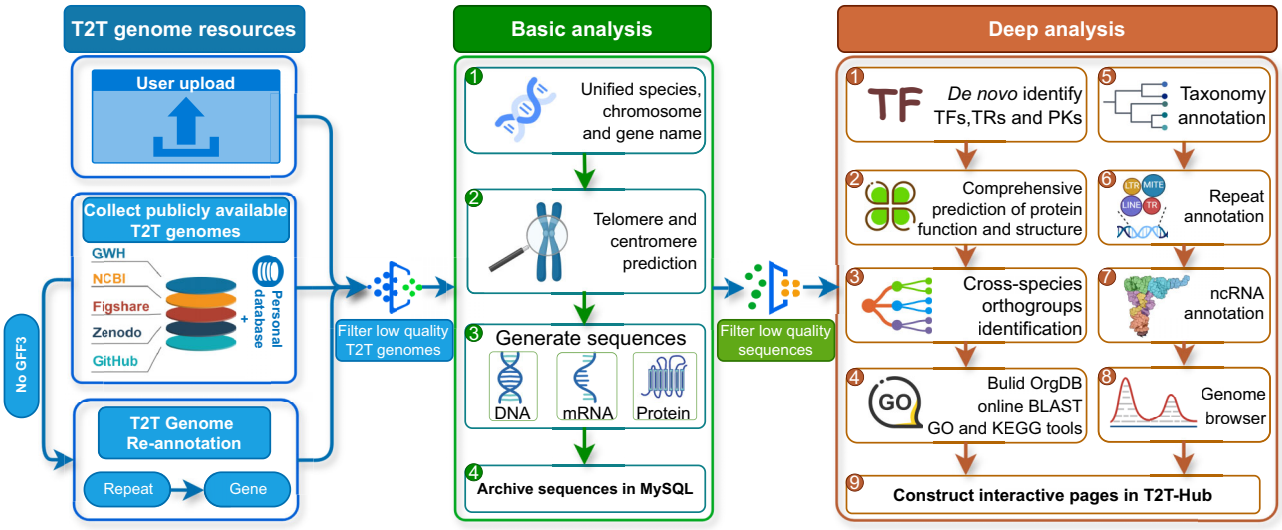
[1]Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[2]Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China
[3]Zhejiang Key Laboratory of Biology and Ecological Regulation of Crop Pathogens and Insects, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China
[4]State Key Laboratory of Vegetation Structure, Function and Construction, Zhejiang University, Hangzhou 310058, China
[5]Zhejiang Key Laboratory of Multi-omics Precision Diagnosis and Treatment of Liver Diseases, Zhejiang University, Hangzhou 310058, China

*To whom correspondence should be addressed. Email: mchen@zju.edu.cn
Correspondence may also be addressed to Dijun Chen. Email: dijun.chen@nju.edu.cn
Correspondence may also be addressed to Cong Feng. Email: ventson@zju.edu.cn
†The first three authors should be regarded as Joint First Authors.

## Abstract

T2T-Hub is a free, web-based platform designed to facilitate the analysis and visualization of telomere-to-telomere (T2T) genomes in plants and animals. The server integrates automated structural and functional annotation workflows with interactive visualization and comparative analysis tools, enabling systematic exploration of complete genome assemblies. T2T-Hub provides a standardized and unified analytical framework specifically tailored for T2T genomes, enabling consistent and comparable analyses across species. To ensure robustness and comparability, we uniformly analyzed 230 high-quality plant and 39 animal T2T genomes using the same standardized workflows, providing users with a consistent reference framework. Users can upload assembled T2T genome sequences together with gene annotation files (GFF3), which are automatically validated and processed through unified analyses including genome quality assessment, telomere and centromere identification, transcription factor prediction, functional annotation, and interactive genome visualization. Importantly, the platform enables interactive genome browsing, searching, and result sharing without requiring any programming skills, thereby lowering the barrier for T2T genome analysis. For internally curated genomes, T2T-Hub additionally provides repeat and noncoding RNA annotation, comparative genomics analyses, and integrated online tools. The platform is freely accessible without login at https://bis.zju.edu.cn/t2thub and https://biobigdata.nju.edu.cn/t2thub.

## Graphical abstract




**Received:** January 4, 2026. **Revised:** April 10, 2026. **Accepted:** April 20, 2026

## Introduction

The advent of third-generation long-read sequencing, especially Oxford Nanopore Technologies ultralong reads, enables telomere-to-telomere (T2T) genome assemblies that are continuous and gap-free [1]. Earlier references remained incomplete, leaving complex regions unresolved, such as GA-rich microsatellites, centromeres, and ribosomal DNA (rDNA) loci and telomeres. Complete T2T assemblies facilitate investigation of genome structure, genetic variation, and centromere evolution, advancing species improvement, disease studies, and basic biology.

Despite many high-quality T2T genomes across plants and animals, challenges remain in analysis, integration, and interpretation. Coordinated workflows for genome quality assessment, annotation, telomere and centromere characterization, and cross-species comparison are typically fragmented and require substantial bioinformatics expertise. Existing repositories, such as Genome Warehouse [2], Ensembl [3], Phytozome [4], and NCBI Genome [5], serve mainly as archives, with limited interactive exploration or T2T-specific analytical workflows.

To address this, we developed T2T–Hub, a freely accessible web platform integrating genome validation, annotation, interactive visualization, gene- and genome-centric search, and comparative analysis modules. Curated high-quality reference genomes demonstrate pipeline robustness and serve as benchmarks. By combining streamlined workflows, comprehensive annotation, and intuitive interfaces, T2T–Hub lowers technical barriers and facilitates functional, evolutionary, and comparative genomics of T2T assemblies.

## Materials and methods

### System overview and job management

T2T-Hub provides an automated analysis pipeline for T2T genome assemblies. After submission of genome sequences and annotations in supported formats, the server assigns a task identifier and generates permanent status and result URLs for monitoring, sharing, and accessing outputs without login (Fig. 1A).

### Input validation and genome quality control

T2T-Hub validates uploaded genome assemblies and annotations prior to analysis (Fig. 1B). Genome sequences must be provided in FASTA format and gene annotations in GFF3 format, with consistent sequence identifiers and valid genomic coordinates.

Genome assemblies undergo quality control to ensure suitability for downstream analyses. A true T2T genome is ideally gap-free and fully assembled; however, many high-quality assemblies may not strictly meet this definition while remaining suitable for analysis. Therefore, T2T-Hub adopts empirically derived thresholds to retain near-T2T assemblies while excluding low-quality genomes.

The total number of gaps must not exceed min($2n$, 50), where $n$ denotes the chromosome number; this threshold was determined based on empirical analysis of 270 T2T or near-T2T genomes and 1354 non-T2T genomes collected from public databases (Supplementary Table S1). The majority of T2T genomes exhibit very low gap counts (with 99.6% having fewer than 50 gaps and 98.9% satisfying the $\leq 2n$ criterion), whereas non-T2T genomes typically show substantially higher gap numbers, demonstrating that this criterion effectively distinguishes high-quality assemblies from fragmented genomes (Supplementary Fig. S1).

To reduce fragmentation, short contigs are filtered according to assembly continuity: Contigs shorter than 10 Mb are removed when N90 exceeds 10 Mb, whereas contigs shorter than 1 Mb are removed when N90 is below 10 Mb. This adaptive strategy is supported by empirical observations showing that contig length strongly correlates with assembly continuity, ensuring stringent filtering for high-quality assemblies while preserving valid sequences in moderately continuous genomes (Supplementary Table S2). Evaluation results indicate that this two-tier strategy retains nearly all sequence content in high-continuity assemblies while effectively removing short, potentially uninformative contigs (Supplementary Fig. S2).

GFF3 annotations must satisfy basic structural constraints, including valid start–end coordinates and strand information ("+" or "−"). Gene features require unique ID attributes, while transcript and subfeature records must contain valid, nonduplicated ID and Parent attributes. Annotation quality is further evaluated at the protein level; if more than 10% of translated proteins contain internal stop codons or lack proper terminal stop codons, the annotation is considered to be of low quality, and the analysis is terminated (Fig. 1B).

Analyses are automatically halted when any validation or quality control criterion is not met. Users are encouraged to perform local prechecks using our genomeCheck script prior to data upload.

### Standardization of genome and gene identifiers

To ensure consistency, T2T-Hub standardizes chromosome and gene identifiers (Fig. 1C). Chromosome names not starting with "Chr/chr" (case-insensitive) are automatically renamed to a standardized "Chr1, Chr2, …" format. Users may optionally retain biologically meaningful chromosome labels (e.g., for polyploid genomes) using concise formats such as "Chr1A" and "Chr1B"; otherwise, default standardized names are applied.

Gene identifiers are reconstructed based on species, cultivar, haplotype, and chromosome information while retaining original source IDs for traceability. Species identifiers are derived from abbreviated Latin names (e.g., *Oryza sativa* is encoded as Osa), cultivar names are retained in full (without special characters), and haplotypes are encoded as H1, H2, etc. Chromosome identifiers are encoded by removing the "Chr" prefix and representing chromosome numbers in a standardized format (e.g., Chr1 is encoded as 01 and Chr1A as 01A), followed by a "G" tag and a six-digit gene number starting from 000010. All components are concatenated using underscores, resulting in identifiers such as "Osa_NIP_H1_01G000010".

### Telomere and centromere identification

Telomeres are identified by scanning chromosome ends for canonical repeats (CCCTAAA in plants and CCCTAA in animals; ≥20 repeats), and centromeres are predicted as tandem repeat-enriched regions (TRRs) after excluding the terminal 100-kb regions to avoid misidentifying telomeric repeats as TRRs (Fig. 1D). Adjacent TRRs separated by less than 50 kb are merged, and the longest TRR is selected as the putative centromere. It should be noted that these predictions are indicative, as centromeres lacking tandem repeat structures re-

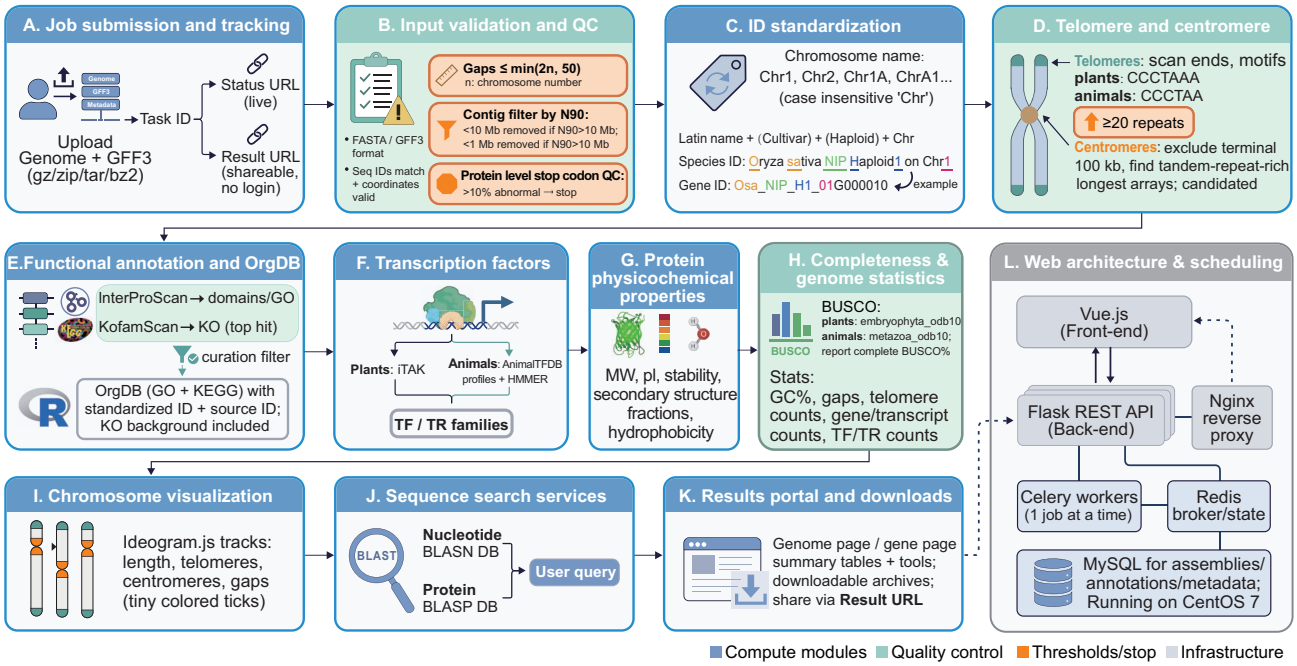



**Figure 1.** Overview of the T2T-Hub workflow and web architecture. (**A**) Job submission and tracking with permanent status and result URLs. (**B**) Input validation and quality control of genome assemblies and annotations. (**C**) Standardization of chromosome and gene identifiers with source ID retention. (**D**) Telomere and centromere detection based on repeat motif scanning and tandem repeat enrichment. (**E**) Functional annotation and construction of species-specific OrgDB. (**F**) Identification of transcription factors and transcriptional regulators. (**G**) Calculation of protein physicochemical properties. (**H**) Genome completeness assessment and summary statistics. (**I**) Chromosome-level visualization of structural features. (**J**) Sequence search services using nucleotide and protein BLAST databases. (**K**) Results portal with genome- and gene-centric pages and downloadable outputs. (**L**) Modular web architecture supporting scalable analysis and online services.

quire experimental validation not supported in the current framework.

## Functional annotation and transcription factor identification

Protein-coding genes are annotated with InterProScan (v5.59-91.0; -f GFF3 –iprlookup –goterms -pa -dp) [6] and KofamScan (v1.3.0; -E 1e-5 –report-unannotated) [7] to assign GO terms and KEGG orthologs (KO), with redundant KO annotations removed by retaining only the highest-scoring assignment per gene, with species-inappropriate pathways removed (Fig. 1E). Curated annotations are integrated into species-specific OrgDBs for enrichment analysis while retaining source identifiers. Transcription factors are identified using iTAK (v1.8; default parameters) (plants) [8] or AnimalTFDB (v4.0)-based HMM searches (animals) [9], and protein physicochemical properties are computed using Biopython (v1.86) ProteinAnalysis module with default parameters (Fig. 1F and G) [10].

## Genome completeness assessment and statistics

Genome completeness is evaluated using BUSCO (v5.8.0) with lineage-specific datasets (embryophyta_odb10 for plants and metazoa_odb10 for animals) [11], and completeness scores are reported based on complete BUSCOs (Fig. 1H). Summary genome statistics, including assembly continuity metrics, GC content, gap distribution, telomere counts, gene and transcript numbers, protein classes, and transcription factor abundance, are calculated and reported.

## Phylogenetic analysis

Users can choose whether to enable phylogenetic analysis on the Analysis module page. If enabled, users may select up to four genomes from T2T-Hub as reference species for OrthoFinder (v3.0.1b1) analysis [12]. It should be noted that enabling this function can substantially increase the analysis time, and users are advised to select it as needed.

## Visualization, data storage, and online services

Chromosome visualization is provided by Ideogram (v1.53.0; https://github.com/eweitz/ideogram/) with integrated telomere, centromere, and gap information (Fig. 1I). All data are stored in MySQL (https://www.mysql.com/) to support querying, with genome-specific BLAST databases and interactive browsing via JBrowse 2 (Fig. 1J) [13, 14]. Results are organized into downloadable archives and dedicated genome and gene pages (Fig. 1K).

## Web server implementation

T2T-Hub is built on a modular architecture with a Vue 2 (https://vuejs.org/) front end and a Flask (https://flask.org.cn/)-based backend providing REST APIs for job submission, monitoring, and result retrieval, with Nginx (https://nginx.org/) as a reverse proxy (Fig. 1L). Asynchronous task execution is handled by Celery (https://github.com/celery/) with Redis (https://redis.io/) as the message broker and state backend to ensure stable scheduling and resource-controlled execution. Genome data, annotations, and metadata are stored in MySQL (https://www.mysql.com/), and the server is deployed on CentOS 7 (https://www.centos.org/) for long-term stability (Fig. 1L).

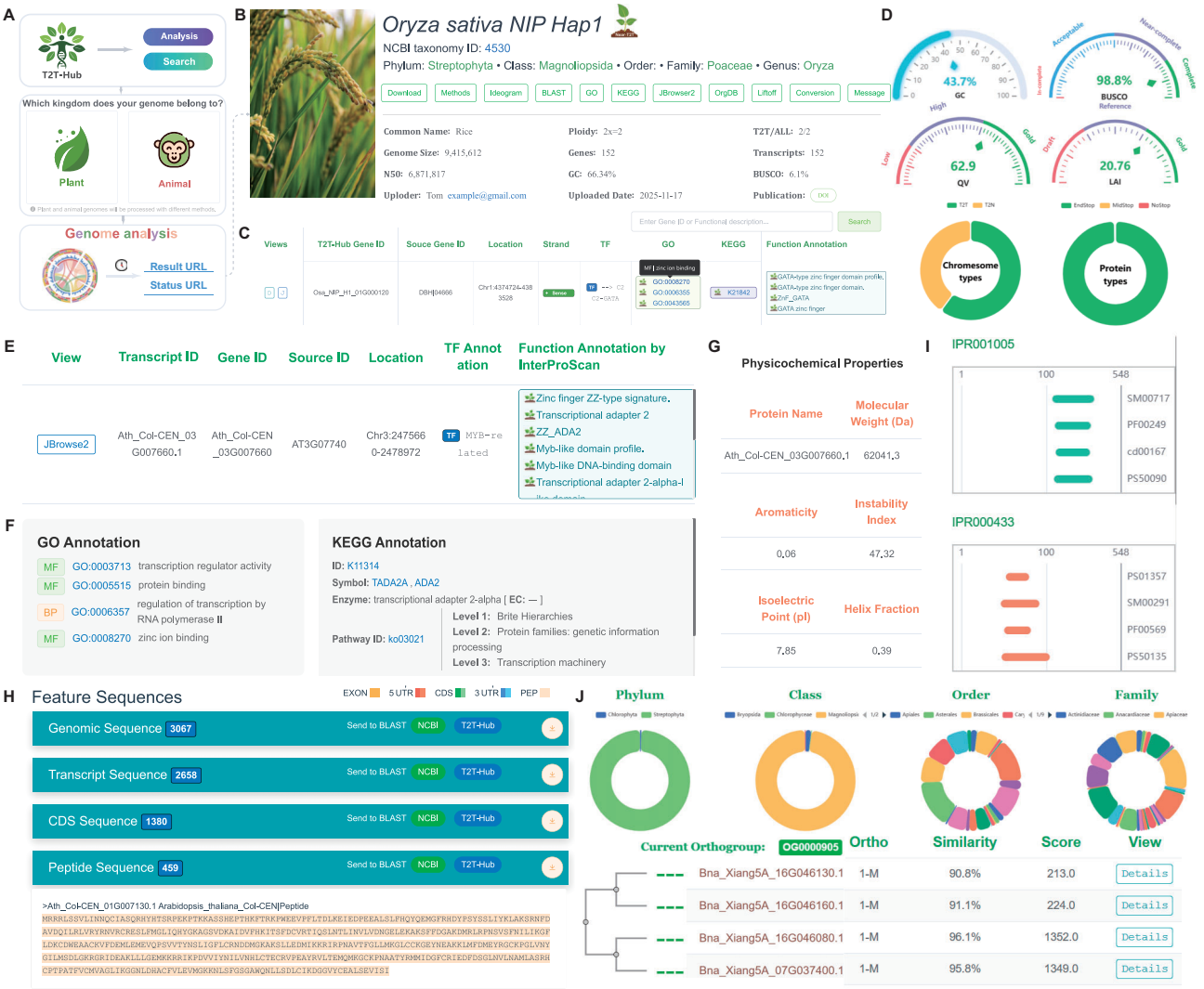



**Figure 2.** Interactive genome and gene page generated by T2T-Hub for a user-submitted T2T genome. (**A**) Simplified workflow for generating a shareable URL to access an interactive genome browser page. (**B**) Genome overview and integrated analysis modules. (**C**) Gene-level exploration table with searchable features. (**D**) Assembly and annotation quality summaries. (**E**) Transcript information and functional annotation. (**F**) Detailed information on GO and KEGG annotations. (**G**) Protein physicochemical properties. (**H**) Multilevel gene and transcript sequences with integrated similarity search tools. (**I**) Conserved protein domain architectures. (**J**) Taxonomic distribution and relationships of homologous genes. This figure illustrates the automatically generated interactive interface following data submission, demonstrating the practical usage of T2T-Hub.

## Results

### Analysis and visualization of user-submitted T2T genomes

For each submitted telomere-to-telomere genome, T2T-Hub generates an interactive genome page integrating assembly statistics, annotation summaries, and downstream analyses (Fig. 2A). The genome page provides an overview of taxonomic information, genome size, GC content, gene and transcript counts, BUSCO completeness, and repeat-related statistics for internally curated genomes (Fig. 2B).

Integrated modules include genome download, assembly methods workflow, ideogram visualization, BLAST sequence search, online GO and KEGG annotation, genome browser, and OrgDB usage and installation, genome coordinate liftover using Liftoff [15], and chromosome and gene ID conversion using Conversion, enabling seamless navigation between genome- and gene-level data (Fig. 2B). Gene exploration is supported through interactive tables that allow queries by identifier, keyword, transcription factor family, chromosomal location, strand, and GO/KEGG annotation, so users can freely search or filter for genes or homologous gene sets of interest (Fig. 2C).

Assembly quality is summarized using intuitive visual metrics, including GC content and BUSCO completeness, both computed automatically by the pipeline without requiring user input; consensus QV and the LTR Assembly Index (LAI, for plant genomes), which are incorporated when provided by users; and genome classification into T2T, T2N, or N2N types based on telomere configuration, irrespective of internal gaps. Protein-coding sequences are further categorized as EndStop, MidStop, or NoStop according to the position of translational stop codons (Fig. 2D). Although T2T-Hub accepts annotations with up to 10% MidStop or NoStop proteins, such cases are not recommended and are highlighted as potential quality concerns. Collectively, these summaries provide an intuitive and unified assessment of genome quality and annotation completeness.

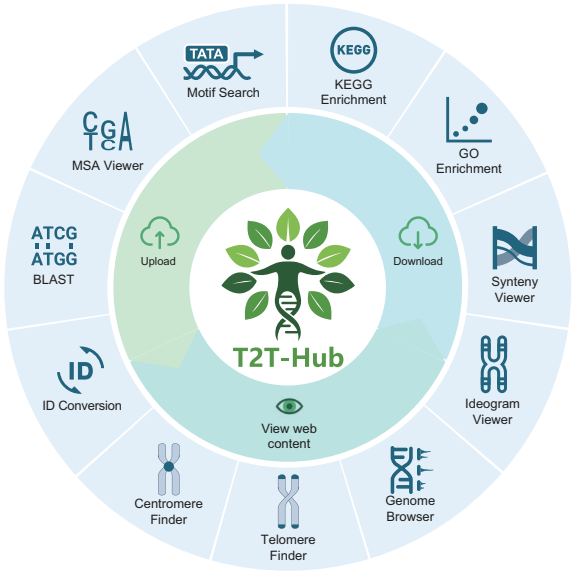



**Figure 3.** Overview of the integrated analysis and visualization tools.

### Analysis and visualization of gene-centric information

For each gene in the uploaded genome, T2T-Hub generates an interactive gene page integrating transcript features, functional annotation, sequences, and comparative analyses. The page summarizes gene and transcript IDs, coordinates, strand, transcription factor classification, and GO, KEGG, and InterProScan annotations (Fig. 2E and F). Protein-coding genes include physicochemical properties such as molecular weight, isoelectric point, aromaticity, and instability index (Fig. 2G). Gene-related sequences (genomic DNA, transcript, CDS, peptide) are downloadable, with links to internal and NCBI BLAST searches (Fig. 2H). Visualization of protein domain architectures was also performed (Fig. 2I).

For curated genomes, homologs are identified and visualized with phylogenetic trees and taxonomic distributions (Fig. 2J). Available gene expression profiles allow exploration of tissue- or condition-specific patterns. The gene page thus provides a concise, comprehensive framework for gene-level exploration.

### Tools for interactive analysis and data exploration

T2T-Hub provides a suite of integrated online tools for both user-submitted and curated genomes (Fig. 3). Sequence similarity is supported via nucleotide and protein BLAST, while genome visualization uses JBrowse2 for dynamic inspection of genes, repeats, centromeres, telomeres, and other tracks.

Functional analysis is enabled through GO and KEGG enrichment using species-specific OrgDBs. Identifier conversion maps T2T-Hub gene IDs to original source IDs. Chromosome-scale features can be explored via interactive ideograms, and telomere and centromere regions through dedicated finder modules.

Comparative genomics is supported with synteny visualization and homologous region exploration. In addition, coordinate translation between genomes is supported through the integration of Liftoff [15], allowing efficient transfer of gene annotations and genomic coordinates between user-submitted and reference genomes.

Additional utilities include primer design, sequence retrieval, motif search, and multiple sequence alignment visualization. All tools are accessible via a unified web interface without login, with results linked to genome and gene pages for traceability.

## Discussion

The rise of T2T genome assemblies offers new opportunities but poses challenges in analysis and reuse. T2T-Hub provides a unified web platform for systematic analysis and visualization of complete genomes, enforcing standardized input validation, quality control, and annotation workflows. Unlike existing repositories, it emphasizes complete assemblies with integrated visualization and functional tools, enabling comparisons against curated high-quality genomes. T2T-Hub links assembly summaries with functional annotation, homolog identification, and expression data, supporting sequence searches, enrichment analyses, structural visualization, and comparative genomics, with results accessible via persistent, shareable URLs.

Current limitations include restricted advanced analyses for user-uploaded genomes and indicative centromere predictions. Future development will expand species coverage, integrate additional omics data, and extend advanced analyses to user submissions while refining centromere and repeat characterization. Overall, T2T-Hub offers a standardized, accessible platform bridging genome assembly and interpretation for functional and comparative genomics in plants and animals.

## Acknowledgements

The authors are grateful to the members of Ming Chen's laboratory and Tongxuan Lv for helpful discussions and valuable comments on Fig. 1.

*Author contributions*: Haoyu Chao (Conceptualization [equal], Formal analysis [equal], Methodology [equal], Software [equal], Validation [equal], Writing – original draft [equal]), Zhonghao Ruan (Formal analysis [equal], Methodology [equal], Software [equal], Validation [equal]), Zhuojin Li (Formal analysis [equal], Investigation [equal], Visualization [equal], Writing – review & editing [equal]), Xinkai Zhou (Software [supporting]), Shilong Zhang (Software [supporting]), Xingxing Shen (Software [supporting], Writing – review & editing [supporting]), Cong Feng (Supervision [equal]), Dijun Chen (Supervision [lead]), and Ming Chen (Conceptualization [lead], Funding acquisition [lead], Project administration [lead], Supervision [lead], Writing – review & editing [equal])

## Supplementary data

Supplementary data is available at NAR online.

## Conflict on interest

The authors declare that there is no conflict of interest associated with this study.

## Funding

This work was supported by the National Natural Science Foundation of China [32270709, 32261133526, 32570787,

32300532, 32070656]; the National Key Research and Development Program of China [2023YFE0112300, 2025YFE0151900]; the 151 Talent Project, and Science and Technology Innovation Leader of Zhejiang Province [2022R52035]; and Jiangsu Collaborative Innovation Center for Modern Crop Production and Collaborative Innovation Center for Modern Crop Production co-sponsored by province and ministry. Funding for open access charge: National Natural Science Foundation of China [32270709, 32261133526].

## Data availability

The web server is available at https://bis.zju.edu.cn/t2thub and https://biobigdata.nju.edu.cn/t2thub. It is completely free and open, and does not require registration. The source code for the workflow is freely available at https://github.com/BioOmics/T2T-Hub and https://doi.org/10.6084/m9.figshare.32064600.

## References

1. Garg V, Bohra A, Mascher M *et al.* Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nat Genet* 2024;56:1788–99. https://doi.org/10.1038/s41588-024-01830-7
2. Chen M, Ma Y, Wu S *et al.* Genome Warehouse: a public repository housing genome-scale data. *Genom Proteom Bioinform* 2021;19:584–9. https://doi.org/10.1016/j.gpb.2021.04.001
3. Harrison PW, Amode MR, Austine-Orimoloye O *et al.* Ensembl 2024. *Nucleic Acids Res* 2024;52:D891–9. https://doi.org/10.1093/nar/gkad1049
4. Goodstein DM, Shu S, Howson R *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 2012;40:D1178–86. https://doi.org/10.1093/nar/gkr944
5. Kitts PA, Church DM, Thibaud-Nissen F *et al.* Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res* 2016;44:D73–80. https://doi.org/10.1093/nar/gkr944
6. Jones P, Binns D, Chang HY *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;30:1236–40. https://doi.org/10.1093/bioinformatics/btu031
7. Aramaki T, Blanc-Mathieu R, Endo H *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 2020;36:2251–2. https://doi.org/10.1093/bioinformatics/btz859
8. Zheng Y, Jiao C, Sun H *et al.* iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant* 2016;9:1667–70. https://doi.org/10.1016/j.molp.2016.09.014
9. Shen WK, Chen SY, Gan ZQ *et al.* AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res* 2023;51:D39–45. https://doi.org/10.1093/nar/gkac907
10. Cock PJ, Antao T, Chang JT *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;25:1422. https://doi.org/10.1093/bioinformatics/btp163
11. Manni M, Berkeley MR, Seppey M *et al.* BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 2021;38:4647–54. https://doi.org/10.1093/molbev/msab199
12. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y
13. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460–1. https://doi.org/10.1093/bioinformatics/btq461
14. Diesh C, Stevens GJ, Xie P *et al.* JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* 2023;24:74. https://doi.org/10.1186/s13059-023-02914-z
15. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics* 2021;37:1639–43. https://doi.org/10.1093/bioinformatics/btaa1016

**Received:** January 4, 2026. **Revised:** April 10, 2026. **Accepted:** April 20, 2026