Journal of Integrative Agriculture 2026, XX(X): XXXX–XXXX

Available online at ScienceDirect

Journal of Integrative Agriculture

REVIEW

Telomere-to-Telomere Genome Assembly and Its Application Progress in Domesticated Animal Genetics Research

Dubala Wu[1], Binhong Wen[2], Ting Wang[1], Yaxuan Jiao[1], Tianwei Zhang[1], Sile Hu[3], Chun Li[1], Tiecheng Wu[4], Baoli Yu[5], Jianghong Wu[1#]

[1] *College of Animal Science and Technology, Inner Mongolia Minzu University, Tongliao 028000, China*
[2] *College of Animal Science, Inner Mongolia Agricultural University, Hohhot 010018, China*
[3] *College of Life Sciences and Food Engineering, Inner Mongolia Minzu University, Tongliao 028000, China*
[4] *Inner Mongolia Academy of Agricultural & Animal Husbandry Sciences, Hohhot 010031, China*
[5] *Hulunbuir Academy of Agricultural and Animal Husbandry Sciences，Hailaer, 021000, China*

Abstract

The primary determinant of genome assembly difficulty is not genome size per se, but rather the repetitive architecture of the genome. Repetitive regions associated with specialized chromosomal structures are substantially longer than the reads generated by short-read sequencing technologies, thereby posing considerable challenges to accurate assembly. Long-read sequencing technologies, however, are capable of producing reads that exceed the length of repetitive sequences, making it possible to construct seamless telomere-to-telomere (T2T) assemblies and thus providing novel strategies for elucidating genome structure and function. T2T assembly is more readily achievable in animals than in polyploid plant genomes, and notable successes have been achieved across a range of domesticated animal species, with the field now entering a phase of rapid expansion. This review describes the methodological foundations of T2T genome assembly and surveys the current research progress of T2T genome research in domesticated animals. It further examines the applications of T2T genomes in domesticated animal genetics research and discusses the associated challenges, with the aim of advancing the use of complete genome assemblies in the study of species evolution, genetic improvement, and related disciplines.

Keywords: Domesticated animal, T2T assembly, assembly methods, applications, challenges

1. Introduction

Genomic sequences represent a critical resource for molecular breeding in animals, encoding the complete genetic information underlying all economically important traits, including growth, reproduction, and disease resistance. Nevertheless, traditional animal reference genomes have long been plagued by numerous assembly gaps, particularly in regions such as telomeres, centromeres, repetitive sequences, and the Y chromosome, which have historically proven resistant to accurate assembly. This persistent incompleteness has rendered vast amounts of genetic variation, functional

Received XXXX; Received in revised form XXXX; Accepted XXXX; Available online XXXX
#Correspondence Jianghong Wu, E-mail: wujianghonglong@126.com

doi: 10.1016/j.jia.2023.10.019

genes, and structural information effectively invisible, thereby impeding downstream analyses and compromising the design and interpretation of experimental studies (Miga *et al*. 2021). Third-generation sequencing (TGS), characterized by its long read lengths and high base-level accuracy, has overcome these longstanding limitations by resolving repetitive regions, improving assembly contiguity and accuracy, deciphering complex structural variants (SVs), and enabling haplotype-resolved phasing, collectively driving the production of high-quality, gap-free genome assemblies (Wang, *et al*. 2022).

T2T genome assembly refers to the complete sequencing and assembly of all chromosomes in an organism, achieving uninterrupted coverage from the telomere at one end of each chromosome to the telomere at the other, with no intervening gaps or only an extremely minimal number of unresolved sites (Varshney *et al*. 2020). T2T genome assembly is enabled by long-read sequencing technologies, specifically Pacific Biosciences (PacBio) High-Fidelity (HiFi) sequencing and Oxford Nanopore Technologies (ONT), which have fundamentally transformed approaches to genome assembly. As demonstrated by recently published high-quality reference genomes for humans and sheep, reads have become progressively longer and more accurate, enabling the resolution of previously uncharacterized genomic regions and addressing many of the limitations inherent in early draft assemblies. Consequently, contiguous sequences (contigs) can be anchored to chromosomes with greater precision, and the number of residual gaps has been substantially reduced (Nurk *et al*. 2022; Luo *et al*. 2025).

Although SNP-based analytical technologies have become highly sophisticated and yield substantial amounts of genetic information, many functional variants that influence complex traits, including insertions and deletions, structural rearrangements, and copy number variations, are frequently located within genomic regions that have been inaccurately assembled or remain unresolved in existing reference genomes (Igolkina *et al*. 2025). Consequently, T2T genome assemblies provide high-quality reference sequences that facilitate the identification of causal genes that could not be accurately localized in previous genome versions. Moreover, a comprehensive characterization of genomic variation within breeding populations and germplasm resources is expected to substantially improve the precision and efficiency of livestock breeding programs (Kalbfleisch *et al*. 2024). Based on recent advances in T2T genome assembly, this review systematically introduces the methodological frameworks underlying T2T genome assembly and provides a comprehensive overview of the current research progress of T2T genome research in domesticated animals. It further discusses the applications of T2T genomes in domesticated animal genetics research along with the challenges that remain to be addressed, thereby offering a theoretical foundation for the advancement of genome-assisted breeding in animal science.

2. T2T genome assembly method

T2T genome assembly relies on third-generation long-read sequencing data, complemented by Hi-C chromatin conformation capture technology and sophisticated bioinformatics tools. This integrated approach enables the construction of a complete genome, transitioning from individual DNA fragments to a gapless, chromosome-level sequence (Fig. 1).

2.1 DNA extraction and library construction

The acquisition of high-molecular-weight (HMW) DNA fragments during T2T genome construction is critical for long-read sequencing. When DNA fragments are too short, they cannot span extensive regions of highly repetitive sequences, preventing assembly software from determining the correct order and orientation of fragments, thereby resulting in persistent gaps. Several methods are commonly employed for DNA extraction, including agarose embedding (Schwartz *et al*. 1984), organic solvent extraction encompassing the CTAB method(Guha *et al*. 2018) and phenol-chloroform extraction(Miller *et al*. 2018), column-based purification (Miller *et al*. 2018), and bead-based purification utilizing the Nanobind CBB Big DNA Kit (Nurk *et al*. 2022) and other commercial high-molecular-weight reagent kits (Bari *et al*.2025; Barton *et al*. 2026) . Among these approaches, the agarose embedding method effectively prevents DNA shearing; however, the procedure is laborious and time-consuming (Schwartz *et al*. 1984). The CTAB method is well suited for plant tissues with high polysaccharide and polyphenol content (Guha *et al*. 2018). Phenol-chloroform extraction offers advantages including low cost and relatively high continuity (approximately 10 kb), but relies on alcohol precipitation, which causes DNA entanglement and shearing, resulting in the loss of ultra-long DNA molecules and N50 values that fail to reach the 100 kb threshold required for T2T sequencing (Miller *et al*. 2018). Column-based purification, which operates through silica membrane or anion exchange columns, is rapid, safe, and highly standardized, but generally yields DNA of comparatively shorter fragment lengths (Miller *et al*. 2018). The Nanobind CBB Big DNA Kit enables the extraction of

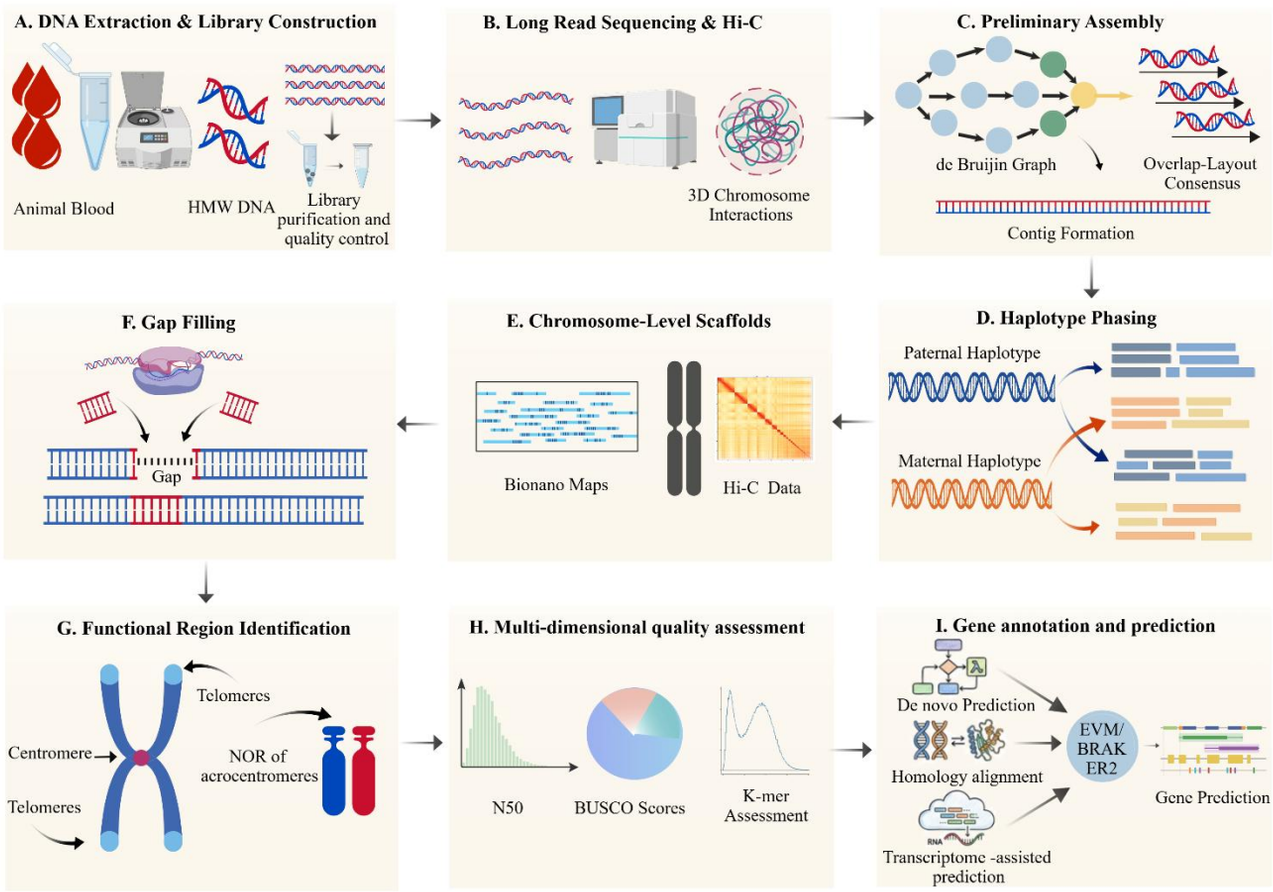

Fig. 1 Schematic diagram of the T2T genome assembly process in domestic animals. A. Animal blood extraction, HMW DNA extraction, and library construction. B. High-precision long-length sequencing and Hi-C chromosome-level assembly. C. Core genome assembly algorithms include two types: overlap-layout consensus and de Bruijn graph. These two methods are combined in T2T assembly of domesticated animals. D. Homologous chromosomes from different parents are isolated to obtain two complete haplotype genomes. E. Chromosome-level genome assemblies in domesticated animals have been achieved through the integrative application of Bionano optical mapping and Hi-C chromatin conformation capture technologies. F. Gap filling is performed in the assembled genome to obtain gap-free complete sequences. G. Centromeres, telomeres, and nucleolar organizing regions (NORs) of chromosomes are identified and validated. H. The assembled genome is then assessed for continuity, integrity, and accuracy. N50 is primarily used to determine genome continuity; a higher N50 indicates better genome continuity. BUSCO is a core tool for assessing genome integrity, while k-mer consistency is the main method for assessing genome accuracy. I. Gene annotation and prediction were performed by integrating de novo prediction, homology alignment, and transcriptome-assisted prediction results using EvidenceModeler (EVM) and BRAKER2. Created with BioGDP. Com.

ultra-long DNA through a workflow involving gentle cell lysis, proteinase K digestion, and purification using nanoscale magnetic disk (Nanobind) technology, collectively producing high-quality, HMW DNA (Nurk *et al*. 2022). Other commercial HMW kits, including the MagAttract HMW DNA Kit, the DNeasy UltraClean Microbial Kit, and the MagMAXTM Microbiome Ultra Nucleic Acid Isolation Kit, are primarily designed for the extraction of bacterial genomic DNA (Bari *et al*.2025; Barton *et al*. 2026).

Given these considerations, DNA extracted using the Nanobind CBB Big DNA Kit is particularly well suited for T2T sequencing of animal genomes. Researchers (Liu *et al*. 2023) employed this kit to perform DNA extraction, subsequently completing the assembly of the yak pangenome. In addition, several best practices are essential throughout the DNA extraction process, including gentle cell lysis, avoidance of vortexing or forceful pipetting, use of wide-bore or trimmed pipette tips, and purification via magnetic beads or dialysis rather than alcohol precipitation (Gershman *et al*. 2022).

2.2 Long read sequencing and Hi-C technology

The advent of third-generation sequencing technologies has ushered in a new era for animal genome sequencing. In 2019, PacBio, a company based in the United States, introduced PacBio HiFi sequencing technology, which achieves high base-level accuracy through circular

consensus sequencing (CCS), attaining accuracy rates exceeding 99%. This technology enables precise error correction at the nucleotide level and facilitates the construction of high-quality contigs, thereby resolving the longstanding trade-off between long read lengths and high error rates that had previously characterized third-generation sequencing platforms (Wenger *et al*. 2019). ONT, a company based in the United Kingdom, offers nanopore sequencing technology capable of generating ultra-long reads, with maximum read lengths reaching 100 to 200 kb. These ultra-long reads can span highly repetitive genomic regions, including centromeric satellite DNA arrays and rDNA clusters, thereby filling assembly gaps that arise from the inability of shorter reads to traverse such complex regions (Wang *et al*. 2021). Therefore, in the context of T2T genome assembly, HiFi sequencing provides high base-level accuracy that enables the resolution of subtle sequence differences between homologous chromosomes, while ONT ultra-long reads traverse longer and more complex repetitive arrays, delineating their boundaries and bridging assembly gaps. The complementary strengths of these two technologies together enable complete genome assembly. Furthermore, Hi-C, a high-throughput chromosome conformation capture technology, was first introduced in 2009 by the research group of Job Dekker as a method for resolving the three-dimensional spatial organization of chromatin within the nucleus. Beyond its applications in characterizing chromatin architecture, Hi-C has been widely adopted to assist in scaffolding genome assemblies to the chromosome level and to investigate gene regulatory mechanisms (Lieberman-Aiden, *et al*. 2009). During the T2T assembly process, Hi-C data enable the accurate integration of otherwise fragmented contigs, which may individually span tens of megabases, into complete chromosomal sequences. This capability is particularly critical for distinguishing between homologous chromosomes through haplotype phasing and for achieving accurate assembly of sex chromosomes.

2.3 Preliminary genome assembly

The core algorithmic frameworks underlying genome assembly are broadly divided into two categories: the Overlap-Layout-Consensus (OLC) approach, implemented in tools such as Hifiasm, Flye, HiCanu, ALGA, and SAVAGE, and the de Bruijn graph approach, employed by tools including Verkko, SOAPdenovo2, Velvet, SPAdes, and ABySS. These two paradigms are each suited to the read length characteristics of different sequencing technologies (Li *et al*. 2024). In the OLC framework, individual sequencing reads serve as nodes in a graph, with edges representing sequence overlaps between reads; the assembler systematically identifies all pairwise overlaps among reads and progressively merges them into longer contigs. In contrast, the de Bruijn graph approach first decomposes sequencing reads into short subsequences of length k, known as k-mers, which serve as nodes in the graph, with edges connecting adjacent k-mers that share an overlap of k minus one base; complete sequences are then reconstructed by traversing paths through this graph (Cheng *et al*. 2024; Turner *et al*. 2018). While de Bruijn graph methods are highly computationally efficient and capable of handling large and complex genomes, they encounter substantial difficulties in resolving long repetitive sequences, which frequently leads to assembly errors and fragmentation. Consequently, for the assembly of longer and higher-quality genomes, OLC-based methods confer a distinct advantage, as their use of full-length reads enables more reliable resolution of repetitive regions and supports the construction of more contiguous and accurate assemblies.

Currently, the two predominant assembly tools capable of simultaneously integrating ONT and HiFi data are Verkko and Hifiasm. Verkko demonstrates robust performance in incorporating ultra-long reads to simplify the assembly graph; however, it typically requires prior error correction of ONT data (for example, using the HERRO tool), involves a more computationally complex pipeline, demands GPU resources, and operates at comparatively slower speeds (Cheng *et al*. 2025). In contrast, the most recent version of Hifiasm supports the concurrent input of HiFi, ONT, and Hi-C data, enabling an integrated workflow that encompasses sequence assembly, resolution of complex genomic regions, and chromosomal scaffolding within a single framework (Li *et al*. 2024).

For homozygous diploid genomes, the simultaneous use of HiFi and ONT data is generally sufficient to produce a high-quality preliminary assembly. A small number of residual gaps may remain in regions such as rDNA arrays (Wang *et al*. 2022), satellite repeat sequences (Itemose, *et al*. 2022), and segmental duplications (Vollger *et al*. 2023); however, these can be resolved through subsequent gap-filling procedures to yield a complete genome assembly. T2T assemblies of domesticated animals, including sheep (Luo *et al*. 2025), pig (Luo *et al*. 2025), goat (Wu *et al*. 2024), and cattle (Pineda *et al*. 2025), have all been accomplished using these two data types in combination. For polyploid and highly heterozygous genomes, HiFi and ONT data alone may be insufficient for accurate chromosomal phasing, particularly in cases where homozygous regions exceed read length or where highly similar segmental duplications are present. Under such circumstances, the introduction of long-range phasing information through supplementary data types, including Hi-C, parental sequencing (trio binning), or Strand-seq, is necessary to resolve phase signals between haplotypes.

In practice, two primary assembly strategies are employed: primary-alternate assembly (Chin *et al*. 2016) and dual-haplotype assembly (Cheng *et al*. 2022; Rautiainen *et al*. 2023). The primary-alternate assembly approach produces a "primary assembly" representing one complete haploid complement and an "alternate assembly" containing the remaining set of allelic sequences, which may be more fragmented. This strategy typically yields longer contigs and is the preferred approach when the objective is to obtain a complete haploid genome representation (Ebert *et al*. 2021; Li *et al*. 2018; Gao *et al*. 2023). A critical post-processing step in primary assembly workflows involves the use of tools such as Purge Haplotigs (Roach *et al*. 2018) and purge_dups (Guan *et al*. 2020), which integrate sequencing depth information with sequence similarity metrics to identify and remove spurious duplicate sequences arising from paralogous repeats or high sequence similarity between haplotypes, thereby improving the overall accuracy and reliability of the final assembly. The dual-haplotype assembly strategy produces a pair of assemblies, each independently representing one of the two constituent haplotypes, and supports assembly-based variant detection as well as the construction of high-quality pangenomes (Garg *et al*. 2024). Nevertheless, accurate haplotype-resolved phasing in regions characterized by high repeat density or extremely low heterozygosity remains a substantial challenge with this approach.

2.4 Haplotype phase separation

Resolving the two complete haplotype genome sequences derived from homologous chromosomes of different parental origins in diploid organisms is critically important for investigations of heterosis and allele-specific expression. Currently, trio binning represents the most accurate phasing strategy available when parental samples can be obtained. This approach utilizes parent-specific k-mers to directly partition offspring long reads into their respective haplotypes of origin, enabling the efficient recovery of two complete haploid assemblies (Koren *et al*. 2018), and has been successfully applied in F1 hybrid Bengal cats, cattle, and mules (Bredemeyer *et al*. 2021; Oppenheimer *et al*. 2021; Jevit *et al*. 2023). Although this method achieves high phasing accuracy, it is entirely dependent on the availability of parental samples and requires sufficient genetic divergence between the two parents, as exemplified by the horse-donkey cross, rendering it unsuitable for inbred or wild-caught specimens. For genomes characterized by high complexity and elevated repeat content, the Verkko hybrid assembly algorithm has emerged as the predominant approach. This method integrates the base-level accuracy of HiFi reads with the ultra-long read lengths of ONT data, resolving heterozygous sites and performing haplotype phasing directly through the assembly graph (Rautiainen *et al*. 2023). The Verkko hybrid assembly algorithm, in combination with subsequent manual curation, has been successfully employed to complete haplotype-resolved T2T genome assemblies of cattle and sheep Y chromosomes (Olagunju *et al*. 2024). This strategy does not require parental information and is capable of handling highly repetitive genomes; however, it entails substantial computational complexity and may introduce phasing errors in regions of low heterozygosity. An additional phasing strategy, Hi-C-assisted chromosome-level phasing, similarly operates independently of parental data. This approach leverages chromatin spatial proximity information to achieve chromosome-scale haplotype resolution. Applications of this method in animal genomics remain relatively limited; however, Researchers (Garg *et al*. 2020) successfully implemented the DipAsm Hi-C-assisted chromosome-level phasing pipeline to achieve haplotype-resolved assemblies for four publicly available human genomes, including PGP1, HG002, NA12878, and HG00733.

For the majority of domesticated animals, and particularly for wild-caught or rare species, parental samples are frequently unavailable. Consequently, single-individual phasing strategies, such as those implemented in Verkko or Hi-C-assisted approaches, offer substantially greater practical applicability. Among these, the Verkko algorithm, which integrates multi-platform sequencing data, has demonstrated considerable potential for generating complete haplotype-resolved assemblies owing to its capacity to directly resolve regions of high heterozygosity and elevated repeat content.

2.5 Chromosomal-level scaffolds construction

With the continued advancement of sequencing technologies, Hi-C and the Irys optical mapping system developed by Bionano Genomics have emerged as pivotal technologies for elevating preliminary genome assemblies to the chromosomal level. These approaches enhance the contiguity and accuracy of genome assemblies by providing long-range linkage information and correcting misjoins within the initial assembly. Hi-C achieves this through formaldehyde-mediated crosslinking of spatially proximate chromatin fragments, followed by restriction enzyme digestion, ligation, and sequencing to generate a genome-wide chromatin interaction matrix. By exploiting the principle that genomic loci residing on the same chromosome exhibit higher interaction frequencies due to spatial proximity, Hi-C data can be used to cluster, order, and orient the sequences from a preliminary assembly, thereby enabling chromosomal-level scaffolding

(Dudchenko *et al*. 2017). Optical mapping, in contrast, operates on the basis of single-molecule physical maps. This technology involves the extraction of ultra-long DNA molecules, which are subjected to site-specific restriction enzyme digestion and fluorescent labeling, followed by linearization through nanochannel arrays, imaging, and reconstruction of the physical restriction site map. The ultra-long read lengths inherent to this platform enable effective traversal of complex repetitive regions, facilitating the ordering and orientation of sequences and the identification and correction of misassembled junctions (Lam *et al*. 2012).

Because Hi-C technology is inherently limited in its ability to correct sequence-level errors, Bionano optical mapping serves as a complementary approach by effectively identifying and resolving misjoins in sequence assemblies, thereby enhancing overall assembly accuracy. Consequently, the integrative application of Bionano optical mapping and Hi-C chromatin conformation capture technologies in the T2T assembly of domesticated mammalian genomes has substantially improved assembly efficiency and facilitated the achievement of complete, gapless assemblies (Luo *et al*. 2025). In this integrated workflow, high-quality contigs are first generated using long-read sequencing platforms such as PacBio or ONT. These contigs are then scaffolded into longer sequences using the long-range, high-resolution physical map information provided by Bionano optical mapping, with modules such as HybridScaffold employed to merge contigs into extended scaffolds while simultaneously identifying and correcting potential misjoins arising from repetitive elements and other assembly artifacts. Subsequently, three-dimensional chromatin interaction data from Hi-C are incorporated to cluster the optically refined scaffolds into chromosomal groups, within which fine-scale ordering and orientation are performed, culminating in the construction of a chromosome-level genome assembly (Dudchenko *et al*. 2017).

2.6 Gap filling

Even following Hi-C scaffolding and manual curation, residual gaps may persist within the genome assembly. These gaps are predominantly localized to highly repetitive regions such as centromeres and telomeres, and represent the final obstacles to achieving a true T2T assembly. Given that gaps vary considerably in size and structural complexity, distinct filling strategies must be applied accordingly.

For small to medium-sized gaps, the most straightforward approach involves screening for ultra-long ONT reads capable of spanning both flanking sequences of the gap and using them directly for gap filling (Li *et al*. 2018). For large gaps, particularly those situated within centromeric regions where densely packed ultra-long tandem repeats create what are effectively "assembly black holes," local assembly methods are required. Specialized modules such as the Nextgraph component of NextDenovo have been developed to perform localized assembly specifically targeting large segmental repeats. To improve the accuracy of local assembly in repetitive regions and reduce false-positive overlaps, k-mer frequency spectra constructed from short-read sequencing data using tools such as Meryl are employed for validation (Oppenheimer *et al*. 2021). Once the target sequences have been assembled, multiple rounds of polishing using HiFi or high-accuracy short-read data are applied to ensure base-level accuracy in the final sequences (Mc Cartney *et al*. 2022). In practice, these strategies are rarely applied in isolation. Rather, they are flexibly combined and iteratively deployed according to the genomic complexity, heterozygosity, repeat content, and available data for the species under investigation. Luo and wu demonstrated this integrative approach in T2T genome projects for sheep and goats, successfully achieving gap-free genome assemblies through the coordinated application of ONT ultra-long reads, HiFi data, and local assembly strategies (Luo *et al*. 2025; Wu *et la*. 2024). In addition, dedicated gap-filling tools including TGS-GapCloser, quarTeT, TRFill, and FGAP provide systematic assistance or automation of the gap-filling process, substantially improving both the efficiency and accuracy of gap closure across diverse assembly contexts (Xu *et al*. 2020; Lin *et al*. 2023; Wen *et al*. 2025; Piro *et al*. 2014).

2.7 Identification and verification of chromosome functional regions

Centromeres, telomeres, and nucleolar organizer regions (NORs) represent three critical functional domains of the chromosome, and the identification and validation of these regions constitute essential steps toward achieving high-quality T2T genome assemblies.

Centromeric regions are characterized by densely packed satellite DNA sequences and transposable elements, rendering them among the most challenging targets in genome assembly. Beyond the preliminary identification and iterative error correction performed during the assembly workflow using approaches such as k-mer frequency analysis, chromatin immunoprecipitation followed by sequencing (ChIP-seq) with antibodies targeting centromere-specific histone H3 (CENH3) has emerged as the most reliable method for delineating centromere locations and boundaries (Garg *et al*. 2024). ChIP-seq directly captures DNA associated with centromere-specific functional proteins, thereby yielding highly credible results. Accordingly, when appropriate

biological material and antibodies are available in sufficient quantity, ChIP-seq is the preferred validation strategy (Garg *et al*. 2024). In studies of the horse genome, the combination of ChIP-seq and k-mer-based analytical approaches enabled the successful identification and characterization of centromeric repeat sequences, as well as the construction of a chromosome-level map of centromere positions (Cappelletti *et al*. 2023). When resources are limited or suitable antibodies are unavailable, investigators rely more heavily on bioinformatic prediction methods as complementary or alternative approaches. These include sequence feature-based strategies, such as the identification of tandem repeats using Tandem Repeats Finder (TRF), and three-dimensional genomic interaction-based methods, such as the analysis of Hi-C interaction decay patterns characteristic of centromeric domains.

Two primary strategies have been developed for telomere identification and assembly: targeted telomere-directed active assembly and whole-genome scanning validation. The selection between these approaches depends on data availability and assembly objectives. In domesticated animals including sheep (Luo *et al*. 2025), pigs (Luo *et al*. 2025), and goats (Wu *et al*. 2024), targeted telomere-directed active assembly has been employed. This strategy involves screening raw HiFi data for telomere-associated sequences, performing local de novo assembly of these sequences, and subsequently integrating the assembled telomeric sequences into the terminal regions of each chromosome. The principal advantage of this approach lies in its capacity to actively close gaps and achieve complete telomere assembly; however, it imposes stringent requirements on sequencing depth and sequence specificity, and involves a comparatively complex analytical workflow. Whole-genome scanning validation, by contrast, employs software tools such as tidk (Brown *et al*. 2025) and quartet (Lin *et al*. 2023) to search for conserved telomeric repeat sequences throughout the assembled genome, including the canonical TTAGGG and CCCTAA repeats characteristic of domesticated animals. In studies of teleost species such as catfish and perch, these tools have been applied to search assembled genome sequences for telomeric signature sequences, thereby confirming assembly completeness (Nguinkal *et al*. 2024; Zhang *et al*. 2025). While this approach is suited for evaluating assembly integrity and verifying the presence of telomeric sequences, it does not enable the active closure of assembly gaps. Accordingly, the attainment of a T2T complete assembly necessitates the adoption of active assembly strategies, whereas genome quality assessment efforts predominantly employ scanning-based validation approaches.

NORs are composed of tandemly repeated rDNA arrays, and their identification relies on a combination of bioinformatic prediction and experimental validation. When previously characterized rDNA sequences are available, homology-based alignment using BLASTN is the method of choice; in the absence of suitable reference sequences, de novo prediction tools such as RNAmmer are employed instead. Nevertheless, precise physical localization of NORs on chromosomes necessitates confirmation through fluorescence in situ hybridization (FISH) (Porto-Foresti *et al*. 2002). Although FISH involves a comparatively lengthy experimental workflow, it provides accurate chromosomal localization of NORs that cannot be achieved by computational methods alone. In practice, bioinformatic approaches are therefore typically applied first for rapid screening and preliminary prediction, followed by FISH-based validation of key findings for final confirmation. This combinatorial strategy has been successfully applied to the characterization of rDNA loci in species including sheep (Luo *et al*. 2025) and pigs (Luo *et al*. 2025), demonstrating its effectiveness as a standard framework for NOR identification in T2T genome projects.

2.8 Multidimensional quality assessment

To provide a reliable sequence foundation for functional gene mining and structural variation analysis, multidimensional quality assessment of the assembled genome is of paramount importance. This evaluation encompasses three key dimensions: contiguity, completeness, and accuracy.

Contiguity assessment relies primarily on two metrics: assembly size and N50 (Li *et al*. 2024). A higher N50 value indicates that a greater proportion of the assembly is represented by longer contiguous sequences, reflecting superior assembly contiguity. The T2T genome assembly of the Minzhu pig published in 2025 achieved an N50 value three times that of the Sscrofa11.1 reference genome, underscoring the exceptional contiguity of its assembled sequences (Zong *et al*. 2025). Tools available for evaluating assembly size and N50 include QUAST (Gurevich *et al*. 2013) and the k-mer-based GenomeScope (Vurture *et al*. 2017), among others. QUAST provides comprehensive quality assessment, including alignment to a reference genome or validation supported by read mapping, while k-mer-based GenomeScope analysis is applied in the early stages of assembly to estimate genomic characteristics and thereby evaluate the plausibility of the resulting assembly size.

The core tool for completeness assessment is BUSCO, which employs a curated set of evolutionarily conserved, typically single-copy orthologous genes as a reference framework. By evaluating the presence status of these genes in the target genome, classifying each as complete, duplicated, fragmented, or missing, BUSCO

quantitatively measures both genomic completeness and redundancy (Manni *et al*. 2021). In the genome assemblies of sheep (Luo *et al*. 2025), pigs (Luo *et al*. 2025), and goats (Wu *et al*. 2024), BUSCO annotation rates exceeded 96%, establishing these assemblies as reliable reference genomes. Although BUSCO enables cross-species comparisons, it is relatively insensitive to non-coding regions such as centromeres. To address this limitation, T2T assemblies have incorporated the analysis of Previously Unresolved Regions (PURs), which specifically evaluates the extent to which complex repetitive regions have been successfully resolved, representing a critical criterion for validating gap-free assemblies. A variety of tools are available for PURs analysis: BEDTools (Quinlan *et al*. 2010) facilitates the association of PURs with gene and repeat sequence annotations, while Mosdepth (Pedersen *et al*. 2018), Gffread (Meng *et al*. 2020), and DupScan (Lu *et al*. 2023) are responsible for sequencing depth coverage evaluation, structural validation, and the characterization of repetitive element distributions, respectively.

Accuracy assessment is conducted at two levels: base-level and structural-level. K-mer concordance evaluation, as implemented in tools such as Merqury, enables the computation of quality value (QV) scores and the detection of phasing errors. QV scores quantify base-level accuracy and are expressed as Phred-scaled quality values, calculated from the proportion of contig k-mers absent from the sequencing reads. Phasing error detection, on the other hand, assesses the accuracy of haplotype phasing by examining whether allelic sequences have been erroneously collapsed or intermingled (Ewing *et al*. 1998; Rhie *et al*. 2020). Both metrics have been widely adopted as validation standards in human genome research (Logsdon *et al*. 2025). Synteny analysis provides an additional approach to structural accuracy verification, achieved by aligning the newly assembled genome against existing reference genomes (Zhou *et al*. 2022). For example, in studies of chromosomal evolution in the order Carnivora, synteny comparisons between the newly assembled giant panda genome and the reference genomes of dog and cat not only confirmed the structural integrity of the assembly but also elucidated chromosomal-level evolutionary patterns (Fan *et al*. 2019). This approach therefore carries dual value, serving simultaneously as a means of evaluating genomic structural accuracy and as an analytical tool for advancing evolutionary research.

2.9 Gene annotation and prediction

Following the generation of a T2T genome assembly, gene annotation and functional prediction are essential downstream steps, with particular emphasis on the annotation of protein-coding genes. Three principal approaches are available for gene prediction: ab initio prediction, homology-based alignment, and transcriptome-guided evidence-based methods (Garg *et al*. 2022). Ab initio prediction employs pre-trained statistical models to infer gene structures without reliance on external data, making it well suited for the discovery of novel genes; however, its accuracy is relatively limited for predicting splice sites and untranslated regions (UTRs). Homology-based alignment leverages high-quality annotation data from closely related species to determine exon boundaries and splice sites through sequence alignment, yielding high annotation quality for conserved genes(Garg *et al*. 2022). Evidence-based methods utilize RNA-seq data for gene prediction and are capable of identifying splice sites and exons while incorporating UTR information; this approach is particularly effective for detecting low-abundance or novel transcripts, though it may fail to capture genes that are not expressed under the specific experimental conditions employed (Garg *et al*. 2024).

Given that each individual method carries inherent limitations, integrative frameworks are routinely adopted in practice. Tools such as EvidenceModeler (EVM) and BRAKER2 combine and weight the results from all three approaches to generate a more accurate and comprehensive non-redundant consensus gene set (Haas *et al*. 2008; Gabriel *et al*. 2024). In the T2T genome studies of the Minzhu pig and Jinhua pig, the application of this multi-evidence integration strategy predicted 25,054 and 23,924 protein-coding genes, respectively; notably, 14 novel genes absent from the Sscrofa11.1 reference genome were identified in the Minzhu pig (Zong *et al*. 2025; Cao *et al*. 2025).

For T2T genome projects in domesticated animals, the multi-evidence integration strategy is particularly well suited. The relatively conserved gene structures of domesticated animals allow homology-based alignment to provide a reliable annotation scaffold, while the complexity of their transcriptional regulation necessitates transcriptomic evidence for refined correction. Furthermore, the completeness of T2T assemblies creates favorable conditions for the ab initio prediction of previously uncharacterized genes.

3. Research progress in T2T Genome Assembly of Domestic Animals

In 2024, the establishment of the Ruminant Telomere-to-Telomere Consortium (RT2T) marked a pivotal milestone, initiating T2T genome assembly efforts for ruminants including cattle and sheep (Kalbfleisch *et al*. 2024). Concurrently, significant breakthroughs were achieved in

the T2T genome assembly of pigs, chickens, and ducks, collectively advancing comparative analyses of chromosomal architecture and genomic evolution across domesticated animals (Table 1).

In sheep, the first T2T sheep genome (T2T-sheep1.0), based on the Hu sheep breed, achieved complete gap-free assembly of 26 autosomes together with the X and Y chromosomes, enabling the characterization of centromeric satellite DNA families SatI, SatII, and SatIII as well as the structural organization of the Y chromosome (Luo *et al*. 2025). The goat T2T-goat1.0 assembly ranks among the first mammalian genomes globally to achieve T2T-level assembly across all 29 autosomes and both sex chromosomes; on this basis, 63,417 SVs and selection signals associated with cashmere fiber traits were identified (Wu *et al*. 2024). In pigs, the Jinhua pig assembly (JH-T2T) achieved gap-free assembly of 20 chromosomes, while the Wuzhishan pig assembly (T2T-pig1.0) provided complete coverage of 18 autosomes and both sex chromosomes, revealing 194.42 Mb of PURs and a distinctive microsatellite architecture within centromeres (Cao *et al*. 2025). The T2T genome assemblies of the Minzhu pig and Rongchang pig (T2T-Mpig1.0 and T2T-RCpig1.0) together enabled the construction of a graph-based pan-genome for pigs (Zong *et al*. 2025). In cattle, the near-T2T assembly of the Mongolian cattle genome spans 3.1 Gb; however, telomeres at both ends were confirmed for only 3 chromosomes (Su *et al*. 2024). In contrast, the Wagyu T2T genome successfully completed T2T-level assembly for five chromosomes, comprising the complete X chromosome and four autosomes, and for the first time revealed that the cattle X chromosome centromere constitutes a natural neocentromere, lacking the canonical satellite repeat sequences characteristic of typical centromeres (Pineda *et al*. 2025). In avian genomics, the T2T genome assembly of the domestic chicken (GGswu1) successfully resolved the sequences of six microchromosomes (dot chromosomes) that were absent from previous reference assemblies. Although diminutive in size, these chromosomes are characterized by high gene density, elevated GC content, and a greater proportion of repetitive sequences, making them critical for understanding the evolution of avian genomes (Huang *et al*. 2023). Meanwhile, the Taihu goose T2T genome assembly spans 1.2 Gb, with 33 of the 38 autosomes assembled without gaps, while the Z and W sex chromosomes retain residual gaps; this assembly currently represents the only publicly available T2T-level genomic dataset for geese (Zhao *et al*. 2024). In addition, Yak (Liu *et al*. 2023) and Tibetan wild ass (Wuyun *et al*. 2026) genomes have also been assembled, though neither has attained T2T-level completion.

In summary, the current landscape of T2T genome assemblies in domesticated animals is characterized by a leading position held by even-toed ungulates, represented by pig, sheep, goats, and select cattle breeds, with Chinese indigenous breeds contributing the vast majority of available datasets. In contrast, significant gaps remain in coverage for equids, poultry species beyond chickens and ducks, and waterfowl other than the Taihu goose.

4. Applications of T2T genome assembly in domesticated animal genetics research

Through the complete resolution of centromeric satellite repeat composition and transposable element invasion patterns, researchers have been able to elucidate the structural differences among centromeres across distinct chromosomes as well as the mechanisms underlying the formation of evolutionary strata. Concurrently, gap-free assembly of the X and Y chromosomes has revealed the structural characteristics of sex chromosome-specific regions and shed light on the evolutionary trajectories of sex-determining genes. Furthermore, T2T reference genomes have enabled the rediscovery of a substantial number of protein-coding genes and functional elements that were previously overlooked in incomplete assemblies. Through precise structural variant detection, large-scale genomic rearrangements, including inversions and translocations, have been closely associated with complex traits. In addition, the high-quality genomic maps provide reduced alignment bias for population genetic analyses, thereby facilitating the reconstruction of fine-scale population divergence histories and gene flow networks. (Fig. 2).

4.1 Insights into centromere structure and evolution

Centromeres evolve rapidly across species and exhibit remarkable diversity, with variation observed even among different populations or individuals within the same species. This accelerated evolution is closely associated with species-specific adaptive processes (Logsdon *et al*. 2024). Studies leveraging complete T2T genome assemblies have revealed that human centromeres are predominantly composed of alpha-satellite DNA, which accounts for 2.8% of the complete T2T-CHM13 human genome (Altemose *et al*. 2022). Similarly, the centromeres of most chicken chromosomes are organized around higher-order repeats (HORs), a structural arrangement analogous to that observed in primates (Huang *et al*. 2023), whereas the zebra finch possesses centromeric satellite sequences composed of short repeats ranging from 1 to 6 base pairs (Formenti *et al*. 2025). These findings demonstrate that fundamental differences in centromeric composition exist

Table 1. T2T Genome Assemblies in Domestic Animals.

Species categories	Scientific name	Sex	Data Strategy	Number of assembled haploid chromosomes/Total number of haploid chromosomes	Centromeres	Telomeres	Genome size (Mb)	References
Livestock	Hu Sheep (*Ovis aries*)	Male	52.0×PacBio HiFi+189×ONT+HiC	26+X and Y/27	28	56	2850	Luo *et al*. (2025)
	Churro sheep (*Ovis aries*)	Male	46×PacBio+106×ONT+15×ONT UL	Y/27	1	2	25.9	Smith *et al*. (2024)
	East Friesian sheep (*Ovis aries*)	Female	49×PacBio HiFi+107×ONT+132×HiC	26+X/27	24	41	2960	You *et al*. (2024)
	Arbas Cashmere Goat (*Capra hircus*)	Male	49.3×PacBio HiFi+114.8×ONT+152.4×HiC	29+X and Y /30	31	40	2860	Wu *et al*. (2024)
	Wagyu×Tuli F1 (*Bos taurus*)	Male	58.1×PacBio HiFi +228.8×ONT + HiC	4+X/30	5	10	3140	Pineda *et al*. (2025)
	Wagyu (*Bos taurus*)	Male	53×PacBio HiFi +101×ONT+21×ONT UL	Y/30	1	2	59.4	Smith *et al*. (2024)
	Mongolian cattle (*Bos taurus*)	Male	36.8×PacBio HiFi +50×ONT+ 130.2×HiC	29+ X and Y/30	—	29	3100	Su *et al*. (2024)
	Wuzhishan pig (*Sus scrofa*)	Male	73.98×PacBio HiFi + 65.31×ONT+100× HiC	18+X and Y /19	20	40	2630	Luo *et al*. (2025)
	Min Pig (*Sus scrofa*)	Male	47.69×PacBio HiFi +148.66×ONT+106.97× HiC	18+X/19	19	38	2660	Zong *et al*. (2025)
	Rongchang pig (*Sus scrofa*)	Male	47.69×PacBio HiFi +148.66×ONT+106.97× HiC	18+X/19	19	38	2680	Zong *et al*. (2025)
	Jinhua pig (*Sus scrofa*)	Male	51.1×PacBio HiFi +136.65×ONT+94× HiC+ 50× WGS	18+X and Y /19	17	33	2690	Cao *et al*. (2025)
	Tibetan wild ass (*Equus kiang*)	Male	58×PacBio HiFi +74×ONT +100× HiC+36×NGS	25+X and Y/26	27	54	2490	Wuyun *et al*. (2026)
Poultry	Skie (*Gallus gallus*)	Female	39×PacBio HiFi + 245×ONT+193× HiC	38+Z and W /39	24	36	1090	Zhao *et al*. (2025)
	Mallard (*Anas platyrhynchos*)	Female	36×PacBio HiFi + 207×ONT+116× HiC	39+Z and W/40	36	36	1220	Zhao *et al*. (2025)
	Huxu (*Gallus gallus*)	Female	52×PacBio HiFi + 80×ONT+ HiC	38+Z and W/39	—	—	1600	Huang *et al*. (2023)
	Taihu goose (*Anser cygnoides*)	Female	60×PacBio HiFi + 44×ONT+135 ×HiC	33+Z and W/39	35	—	1200	Zhao *et al*. (2024)
	zebra finch (*Taeniopygia guttata*)	Female	122×PacBio HiFi + 117×ONT+255×ONT UL +57 ×HiC	40/40	40	80	1140	Formenti *et al*. (2025)

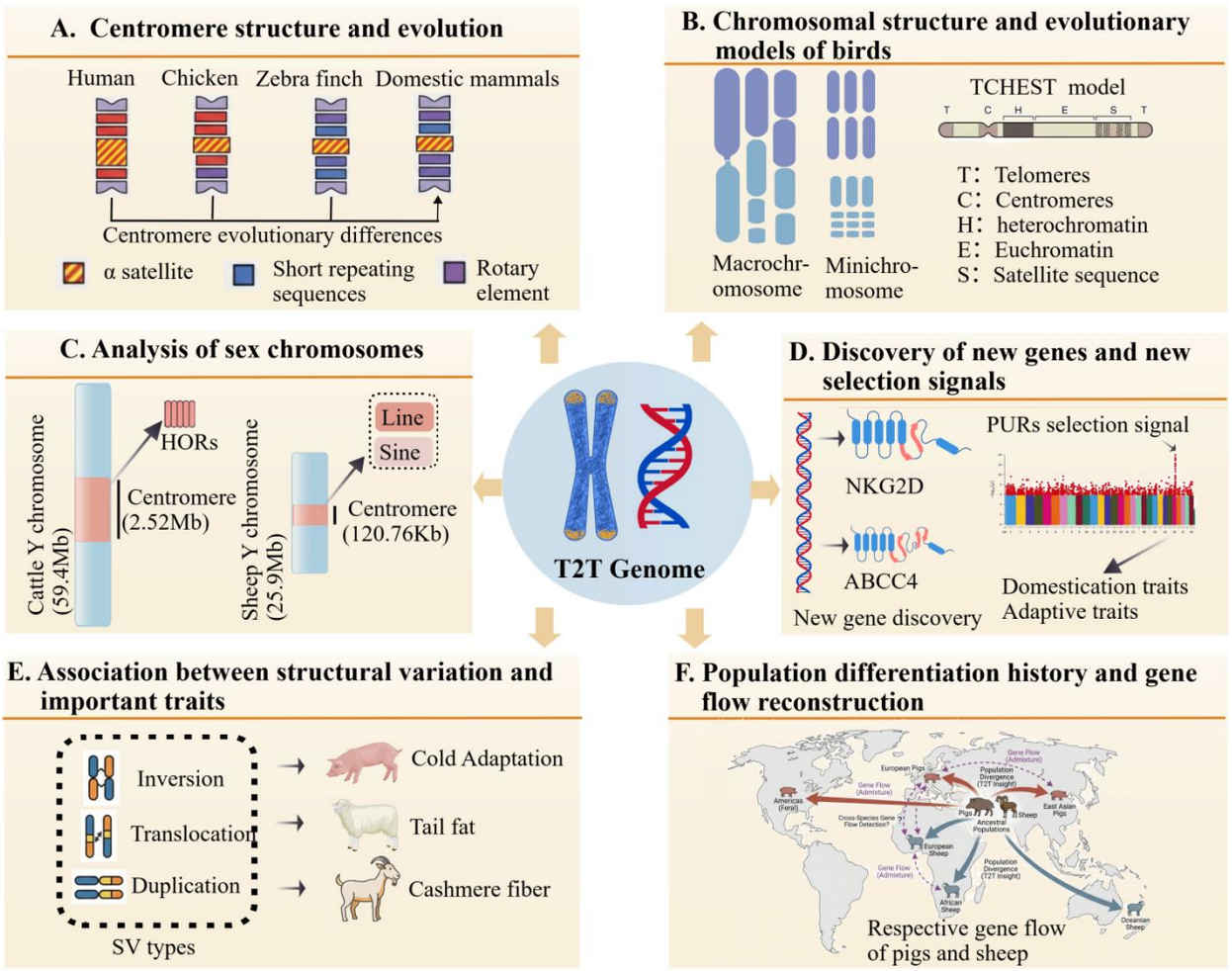

Fig. 2 Applications of T2T genome assembly in domesticated animal genetics research. A. Centromere structure and evolutionary analysis. Human and chicken centromeres are composed of α-satellites and HORs, zebra finches' centromeres are composed of segmental repetitive sequences, while the centromeres of domesticated mammals are rich in transposon elements. B. T2T was able to analyze the macrochromosomes and microchromosomes of domestic chickens and reconstruct the TCHEST pattern of zebra finch chromosomes. C. Differences in Y chromosome size, centromere size, and centromere composition between cattle and sheep. D. Discovery of new genes (NKG2D and ABCC4) deleted or fragmented in the old reference genome and new trait-related selection signals. E. Numerous SVs, including invertions, translocations, and duplications, have been identified in association with economically and biologically important traits in domesticated animals, such as cold adaptation, fat-tail deposition, and cashmere fiber production. F. Discover the history of population differentiation and gene introgression events in pigs and sheep. Created with BioGDP. com

across major taxonomic groups.

In multiple domesticated mammalian species, functional centromeres can be predominantly constituted by transposable elements, including long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and long terminal repeat retrotransposons (LTRs), rather than by satellite DNA. T2T sequencing enables genome-wide quantification and precise localization of transposable elements, facilitating the resolution of their complete sequences, abundance distributions, and evolutionary histories (Tsukahara *et al*. 2025). Studies in sheep (Luo *et al*. 2025), pigs (Luo *et al*. 2025), and goats (Wu *et al*. 2024) have each characterized the genome-wide distribution, content, and region-specific roles of LINEs, SINEs, and LTRs. Further investigation has established that these transposable elements participate in the structural composition of centromeric regions, with LTR elements exhibiting particularly pronounced enrichment and recent insertion activity within centromeres. Notably, the mean insertion age of LTR elements at pig centromeres was estimated at 12.31 million years, substantially more recent than the genome-wide average of 41.11 million years. Comparative analysis with other large mammals, including humans, goats, and sheep, further revealed that pig centromeres exhibit a remarkably low level of invasion by young LTR elements (Luo *et al*. 2025). This finding indicates that the composition and evolutionary trajectory of pig centromeres differ substantially from those of sheep and goats, and that pigs employ a distinctly unique strategy for centromere

construction and maintenance. The species-specificity of centromere composition, on the one hand, reflects the unique selective pressures faced by different lineages during the course of evolution; on the other hand, it suggests that the maintenance of centromere function may rely more heavily on epigenetic mechanisms than on specific DNA sequences.

4.2 Reconstructing the macroscopic structure and evolutionary model of bird chromosomes

Microchromosomes are a pervasive feature of avian genomes, and the 10 dot chromosomes of the domestic chicken are particularly distinctive. These chromosomes are characterized by exceptionally high GC content and elevated DNA methylation levels, unusually large pericentromeric heterochromatic regions, and continuous expansion driven by the proliferation of satellite DNA and testis-specific expressed genes, collectively reflecting unique structural and epigenetic properties (Zhao *et al*. 2025). Through the application of T2T sequencing technology, researchers achieved complete telomeric closure of all autosomes in the domestic chicken — encompassing macrochromosomes, microchromosomes, and dot chromosomes. The resulting assembly spans 50.4 Mb more than the previous reference genome, GRCg6a. Notably, this work marks the first time that six long-elusive small chromosomes (chr29 and chr34–38) have been fully assembled, closing a significant gap in avian genomics (Huang *et al*. 2023). In the zebra finch, T2T assembly not only enabled the complete resolution of all microchromosomes but also led to the proposal of a conserved TCHEST organizational model, describing a canonical chromosomal architecture proceeding from telomere through centromere, heterochromatin, euchromatin, and satellite sequences to the opposing telomere. This framework fundamentally challenges prevailing assumptions regarding centromeric composition and chromosomal evolutionary patterns in birds. The discovery supports the view that avian microchromosomes are highly conserved in both sequence and structure, and traces their origin to an ancestral avian-reptilian karyotype approximately 300 million years ago (Formenti *et al*. 2025). These findings indicate that the canonical avian-like karyotype, comprising 10 pairs of macrochromosomes and approximately 30 pairs of microchromosomes, was already established in diapsid ancestors and has been retained across the majority of modern avian lineages.

4.3 Analysis of sex chromosomes

T2T sequencing has confirmed that sex chromosomes, and the Y chromosome in particular, carry a substantially greater repetitive sequence burden than autosomes. The centromere of the cattle Y chromosome is composed of a 2.5 Mb HORs array assembled from 73 bp monomeric units, whereas the sheep Y centromere spans only 120.7 kb and is constituted by LINE and SINE transposable elements (Olagunju *et al*. 2024). With respect to the X chromosome, transposable elements account for as much as 72.2% of the cattle X chromosome centromere, which additionally exhibits distinctive characteristics including low CENP-A signal intensity, reduced methylation levels, and diminished CpG content. These features collectively demonstrate that transposable elements can serve as core constituents of functionally critical regions on sex chromosomes (Pineda *et al*. 2025). By contrast, the centromeric region of the sheep X chromosome is predominantly organized around satellite DNA composed of HORs, with the satellite repeats classified into three categories: SatI (816 bp), SatII (702 bp), and SatIII (22 bp), the last of which was identified as a novel centromeric unit by Luo and colleagues (Luo *et al*. 2025). Furthermore, the centromeric region of the sheep X chromosome is highly methylated, a pattern resembling that observed in human chromosomes (Luo *et al*. 2025). Taken together, these findings demonstrate that the sex chromosome centromeres of cattle and sheep differ markedly in repeat unit type, size, and epigenetic characteristics, indicating that sex chromosomes have undergone independent and accelerated evolutionary trajectories in each lineage.

T2T assembly has further revealed that the expansion or contraction of amplicon gene families constitutes the primary force driving diversification in Y chromosome size and gene content. The cattle Y chromosome (59.4 Mb) is more than twice the length of its sheep counterpart (25.9 Mb), a disparity attributable predominantly to amplicon region expansion in cattle and pseudogenization in sheep (Olagunju *et al*. 2024). The Y chromosome centromere has also evolved rapidly: the cattle Y centromere (2.52 Mb) is approximately 20 times larger than that of sheep (120.76 kb) (Olagunju *et al*. 2024). These observations indicate that even among closely related species such as cattle and sheep, the length, centromere size, and gene repertoire of the Y chromosome can differ dramatically, thereby serving as a potent driver of rapid evolutionary divergence.

4.4 Discovery of new genes and novel selection signals

By uncovering PURs, T2T assembly enables the discovery of a substantial number of protein-coding genes that were entirely absent from previous reference genomes, thereby greatly expanding our understanding of the functional evolution of the genome. For example, immune-related gene families such as NKG2D and metabolism-associated gene families such as ABCC4, which were absent or

fragmented in goat older reference assemblies, were correctly resolved in T2T assemblies as multiple tandemly arrayed copies (with NKG2D expanding from 10 to 18 copies and ABCC4 fully assembled into 14 tandem repeats), revealing the potential regulatory role of gene dosage in complex trait determination (Wu *et al*. 2024). With respect to the identification of selection signals, a large number of signals associated with domestication, environmental adaptation, and economically important traits were identified within PURs regions that had been entirely overlooked in older references due to sequence absence or misassembly. For instance, 406 novel selection signals covering 56 genes were identified in goat PURs, while key variants associated with domestication (ABCC4) and wool-related traits (FOXQ1) were characterized in sheep (Luo *et al*. 2025; Wu *et al*. 2024). Additionally, the pig Y chromosome harbors two copies of SRY, whereas sheep and goats each carry only a single copy (Luo *et al*. 2025), further underscoring the capacity of T2T technology to present large numbers of functional genes and selection signals in their complete form for the first time. Collectively, these advances provide entirely new targets for pinpointing key genes, dissecting the genetic architecture underlying complex traits, and advancing molecularly guided breeding strategies.

4.5 Detection of structural variations and association with important traits

SVs are genomic sequence rearrangements of 50 bp or greater in length, encompassing deletions, duplications, inversions, translocations, and related classes of variation (Alkan *et al*. 2011). Genomic SVs may exert phenotypic effects through molecular mechanisms including the modulation of gene expression, and complete T2T-level assemblies enable the systematic investigation of structural variation and its associations with economically important traits (Nguyen *et al*. 2023). For example, a study of Minzhu pigs identified nearly 200,000 high-confidence SVs, of which 17 were directly associated with cold adaptation genes; the key regulatory factor TPT1 was confirmed as an important target for cold-tolerant breeding (Zong *et al*. 2025). In the context of sheep tail fat deposition, more than 98% of the flanking SNPs associated with relevant SVs exhibited FST values below the 90th percentile, rendering these variants undetectable by conventional SNP-based selective sweep analyses (Li *et al*. 2023). In goats, SVs associated with cashmere fiber traits involve genes including FGFBP1, KAP9-2, and COL1A1, whereas structural variants linked to domestication selection implicate genes such as ABCC4, MUC6, and SPAG16 (Wu *et al*. 2024). These findings collectively underscore the unique and critical value of SVs analysis in elucidating the genetic basis of economically important traits.

4.6 Reconstruction of population divergence history and gene flow

Complete T2T genome assemblies have revealed novel SNPs and SVs that account for more than 15% of total genomic variation, enabling the construction of more accurate phylogenetic relationships, refining population divergence time estimates to the million-year scale, and facilitating the clear reconstruction of complex introgression histories (Lalli *et al*. 2025). For example, a T2T genomic study of pigs determined that Wuzhishan pigs and Duroc pigs diverged approximately 2.1 million years ago. Furthermore, hundreds of thousands of previously unidentified SNPs were discovered in pericentromeric regions, through which a European introgression event in northeastern Asian domestic pig populations was identified, shedding light on the genetic structure of domestic pig populations and the introgression relationships among Asian pig groups (Luo *et al*. 2025). Similarly, T2T-based research in sheep revealed consistent patterns of divergence and introgression between six geographically defined groups of domestic sheep, representing Central and East Asian, South and Southeast Asian, Middle Eastern, African, European, and American populations, and seven wild sheep species, including European mouflon、Asiatic mouflon、urial、argali、snow sheep、thinhorn sheep and bighorn sheep (Luo *et al*. 2025). Taken together, these cases demonstrate the substantial value of T2T technology in reconstructing species evolutionary histories and tracking gene flow across populations and lineages.

5. Challenges in T2T assembly of domesticated animals

T2T sequencing technology has achieved landmark breakthroughs in domesticated animals, including sheep, goats, and swine. Notably, the research group led by Li Menghua completed the first goat T2T genome assembly incorporating complete Y chromosome and centromeric regions (T2T-goat1.0), as well as the sheep T2T genome assembly (T2T-sheep1.0), with a porcine T2T genome subsequently reported as well. Nevertheless, systematically advancing T2T genome assembly across domesticated animal species at a theoretical level remains confronted by numerous challenges.

5.1 Topological uncertainty and latent mis-assembly risks in highly repetitive intervals

The core challenge in advancing T2T genome assembly for domesticated animals does not lie in whether sequencing reads can be connected end to end, but rather in the fundamental complexity posed by highly repetitive genomic intervals, including centromeric satellite arrays, peritelomeric regions, rDNA clusters, and transposable elements such as LINEs, SINEs, and endogenous retroviruses. Even when these intervals are spanned by long reads, the near-identical k-mer composition within repetitive units makes it exceedingly difficult to establish unique anchoring paths for contigs across extensive repetitive sequences. As a result, contigs lose unique landmark resolution within local regions, and current strategies typically rely on spatial constraint information derived from extremely long reads, Hi-C chromatin conformation capture, and Bionano optical mapping to enforce gap closure. This process, however, converts a portion of topological ambiguity into latent misassembly risk (Luo *et al*. 2025; Pineda *et al*. 2025). In the T2T pig genome, five centromeric satellite repeat units, designated SAT1 through SAT3, SatY1, and SatY2, exhibit complex chromosomal distribution biases and acrocentric chromosome-specific organizational patterns (Luo *et al*. 2025). This finding suggests that when subfamilial differentiation gradients or species-specific expansion and contraction events exist within repetitive sequences, the concealed nature of misassembly errors becomes considerably more pronounced.

5.2 Attenuation of breeding utility attributable to haplotype heterozygosity and reference bias

Domesticated animals have generally undergone prolonged artificial selection and inbreeding, resulting in pronounced breed structure and distinctive haplotype heterozygosity. A substantial proportion of key loci associated with economically important traits are embedded precisely within SVs. If T2T assembly provides only a single reference sequence collapsed from the haplotypes of a single individual, allelic diversity becomes compressed into a single path, introducing systematic reference bias. Consequently, population-specific copy number polymorphisms and presence-absence variants are liable to be misclassified as sequencing gaps or low-quality regions, ultimately undermining the predictive power of downstream genome-wide association studies and genomic selection analyses (Crysnanto *et al*. 2020). The T2T study of the Minzhu pig addressed this limitation by constructing a pan-genome, through which nearly 200,000 high-confidence SVs were identified, 17 of which were directly associated with cold adaptation (Zong *et al*. 2025). Taken together, these findings indicate that T2T assembly must be integrated within a pan-genome framework in order to translate the added resolution into genuine breeding utility.

5.3 Chimeric sensitivity and versioning dilemmas of sex chromosomes

The Y chromosome in mammals and the W chromosome in birds are characterized by extreme heterochromatinization, suppressed recombination, and repeat-driven sequence expansion. Even when gap-free assembly is achieved through ultralong reads and multi-platform sequencing data, boundary definition remains susceptible to read source contamination, including maternal DNA contamination, somatic mosaicism, and tissue-level chimerism, as well as diploid phasing discontinuities (Rhie *et al*. 2023; Smeds *et al*. 2015). Furthermore, the centromeric core alpha-satellite higher-order arrays, rDNA tandem repeat clusters, and amplicon regions of the Y chromosome, as a representative sex chromosome, harbor intrinsic intraspecific copy number fluctuations, subtype chimerism, epigenetic modification polymorphisms, and structural rearrangement polymorphisms (Rhie *et al*. 2023). This complexity implies that a T2T reference for sex chromosomes must be accompanied by population-level multi-version genomic resources in order to genuinely support functional research and practical breeding applications.

5.4 Escalating validation demands in microchromosomes and high gene density regions

In avian species such as chickens and ducks, microchromosomes are characterized by a dual nature of high gene density and extensive transposable element infiltration. Even when gap-free assembly is achieved at the sequence level through T2T technology, the resulting continuity is more prone to becoming continuity that is assembled but increasingly difficult to verify. The burden of validation therefore shifts from conventional assembly metrics, including N50, BUSCO completeness, and alignment rates, to orthogonal breakpoint reproducibility, long-range structural consistency, and functional validation (Zhao *et al*. 2025; Li *et al*. 2022). Human T2T research has previously demonstrated that newly assembled segmental duplication regions corrected thousands of structural errors present in GRCh38, and without cross-platform validation, the assembly quality of such complex repetitive regions would be exceedingly difficult to guarantee (Nurk *et al*. 2022). By analogy, the assembly of avian microchromosomes confronts challenges of a comparable nature and magnitude.

5.5 Nonlinear cost escalation driven by large genome size and transposable element burden

Mammalian species such as cattle, horses, and camels possess substantially larger genomes, ranging approximately from 2.7 to 3.1 Gb, with correspondingly high transposable element content and large proportions of heterochromatin. These characteristics collectively cause the required sequencing depth for ultralong reads, Hi-C resolution, and Bionano optical mapping coverage to escalate in a nonlinear fashion. The scarcity of reliable truth sets further compounds both wet laboratory and computational costs (Formenti *et al*. 2025; Gao *et al*. 2024). Achieving broad cross-species adoption of T2T assembly therefore necessitates systematic investment and the establishment of standardized quality assessment frameworks.

6. Future perspectives

Although T2T assembly of domesticated animals presents various challenges, T2T genomic research is evolving toward a more efficient, integrated, and systematic direction. The democratization and automation of sequencing technology represent a primary trend, exemplified by emerging algorithms such as hifiasm, which hold promise for achieving near-T2T assembly quality from single-platform ONT long-read data alone, dramatically reducing costs and rendering large-scale population-level T2T sequencing a practical possibility. This trajectory will lay the foundation for the construction of T2T super pan-genomes that integrate hundreds of complete genomes within a species, generating unprecedented genetic variation maps and illuminating the key mechanisms underlying adaptation, domestication, and selective breeding. Concurrently, complete T2T sequences provide a structural framework for multi-omics integration, enabling transcriptomic, epigenomic, metabolomic, and single-cell datasets to be precisely anchored to the reference, thereby systematically establishing complete causal pathways from DNA sequence to complex phenotype.

Acknowledgements

This work was supported by the Science and Technology Plan Project in the Inner Mongolia Autonomous Region (2023KJHZ0007, 2025YFDZ0135 and 2023YFDZ0057), National Natural Science Foundation (32260812), Inner Mongolia Natural Science Foundation project (2024MS03077), Inner Mongolia Beef Cattle Breeding Technology Innovation and High-Efficiency Breeding System Construction Project (YZ2023010), and the Central government-guided local science and technology development fund projects (2024ZY0102).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that they did not use AI in the preparation and writing of this manuscript.

Ethical statements

Not applicable.

References

Alkan C, Coe B P, Eichler E E. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, **12**, 363-376.

Altemose N, Logsdon G A, Bzikadze A V, Sidhwani P, Langley S A, Caldas G V, Hoyt S J, Uralsky L, Ryabov F D, Shew C J, Sauria M E G, Borchers M, Gershman A, Mikheenko A, Shepelev V A, Dvorkina T, Kunyavskaya O, Vollger M R, Rhie A, McCartney A M, Asri M, Lorig-Roach R, Shafin K, Lucas J K, Aganezov S, Olson D, de Lima L G, Potapova T, Hartley G A, Haukness M, Kerpedjiev P, Gusev F, Tigyi K, Brooks S, Young A, Nurk S, Koren S, Salama S R, Paten B, Rogaev E I, Streets A, Karpen G H, Dernburg A F, Sullivan B A, Straight A F, Wheeler T J, Gerton J L, Eichler E E, Phillippy A M, Timp W, Dennis M Y, O'Neill R J, Zook J M, Schatz M C, Pevzner P A, Diekhans M, Langley C H, Alexandrov I A, Miga K H. 2022. Complete genomic and epigenetic maps of human centromeres. *Science*, **376**, eabl4178.

Bari A K, Xavier B B, Severs T, Sinha B, Rossen J W A. 2025. Comparison of three commercial DNA extraction kits and assemblers for AMR determinant detection in Pseudomonas aeruginosa and Enterobacter cloacae using long-read sequencing. *Journal of Microbiological Methods*, **239**, 107317.

Barton X W, S Tobe S, Fontaine J B, Oskam C L. 2026. A framework for optimising arthropod DNA quality and quantity for modern sequencing tools using hard ticks (Ixodidae). *Parasitology Research*, **125**, 5.

Bredemeyer K R, Harris A J, Li G, Zhao L, Foley N M, Roelke-Parker M, O'Brien S J, Lyons L A, Warren W C, Murphy W J. 2021. Ultracontinuous Single Haplotype Genome Assemblies for the Domestic Cat (Felis catus) and Asian Leopard Cat (Prionailurus bengalensis). *Journal of Heredity*, **112**, 165-173.

Brown M R, Manuel Gonzalez de La Rosa P, Blaxter M. 2025. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*, **41**, btaf049.

Cao C, Miao J, Xie Q, Sun J, Cheng H, Zhang Z, Wu F, Liu S, Ye X, Gong H, Zhang Z, Wang Q, Pan Y, Wang Z. 2025.

A near telomere-to-telomere genome assembly of the Jinhua pig: enabling more accurate genetic research. *GigaScience*, **14**, giaf048.

Cappelletti E, Piras F M, Sola L, Santagostino M, Petersen J L, Bellone R R, Finno C J, Peng S, Kalbfleisch T S, Bailey E, Nergadze S G, Giulotto E. 2023. The localization of centromere protein A is conserved among tissues. *Communications Biology*, **6**, 963

Cheng H, Asri M, Lucas J, Koren S, Li H. 2024. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nature Methods*, **21**, 967–970.

Cheng H, Jarvis E D, Fedrigo O, Koepfli K P, Urban L, Gemmell N J, Li H. 2022. Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology*, **40**,1332-1335

Cheng H, Qu H, McKenzie S, Lawrence KR, Windsor R, Vella M, Park PJ, Li H. 2025. Efficient near telomere-to-telomere assembly of Nanopore Simplex reads. *bioRxiv* [Preprint], 648685.

Chin C S, Peluso P, Sedlazeck F J, Nattestad M, Concepcion G T, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer G R, Delledonne M, Luo C, Ecker J R, Cantu D, Rank D R, Schatz M C. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, **13**, 1050-1054.

Crysnanto D, Pausch H. 2020. Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery. *Genome Biology*, **21**, 184.

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, Aiden E L. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**, 92–95.

Ebert P, Audano P A, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder M J, Sulovari A, Ebler J, Zhou W, Serra Mari R, Yilmaz F, Zhao X, Hsieh P, Lee J, Kumar S, Lin J, Rausch T, Chen Y, Ren J, Santamarina M, Höps W, Ashraf H, Chuang NT, Yang X, Munson K M, Lewis A P, Fairley S, Tallon L J, Clarke W E, Basile A O, Byrska-Bishop M, Corvelo A, Evani US, Lu T Y, Chaisson M J P, Chen J, Li C, Brand H, Wenger A M, Ghareghani M, Harvey W T, Raeder B, Hasenfeld P, Regier A A, Abel H J, Hall I M, Flicek P, Stegle O, Gerstein M B, Tubio J M C, Mu Z, Li Y I, Shi X, Hastie A R, Ye K, Chong Z, Sanders A D, Zody M C, Talkowski M E, Mills R E, Devine S E, Lee C, Korbel J O, Marschall T, Eichler E E.2023. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Nature Biotechnology*, **41**,1474-1482.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, **8**, 186-194.

Fan H, Wu Q, Wei F, Yang F, Ng B L, Hu Y. 2019. Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. *Genome Biology*, **20**, 267.

Formenti G, Jain N, Medico J A, Sollitto M, Antipov D, Barcellos S, Biegler M, Borges I, Chang J K, Chen Y, Cheng H, Conceição H, Davenport M, De Oliveira L, Duarte E, Durham G, Fenn J, Forde N, Galante P A, Gerhardt K, Giani AM, Giunta S, Kim J, Komissarov A, Koo B, Koren S, Larkin D, Lee C, Li H, Makova K, Masterson P, Murphy T, McCaffrey K, Mercuri RLV, Na Y, O'Connell M J, Ou S, Phillippy A, Popova M, Rhie A, Ruiz-Ruano F J, Secomandi S, Smeds L, Suh A, Tilley T, Vontzou N, Waters P D, Balacco J, Jarvis E D. 2025. The complete genome of a songbird. *bioRxiv* [Preprint],682431.

Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *bioRxiv*,**10**,544449.

Gao Z, Lu Y, Chong Y, Li M, Hong J, Wu J, Wu D, Xi D, Deng W. 2024. Beef Cattle Genome Project: Advances in Genome Sequencing, Assembly, and Functional Genes Discovery. *International Journal of Molecular Sciences*, **25**, 7147.

Garg S, Fungtammasan A, Carroll A, Chou M, Schmitt A, Zhou X, Mac S, Peluso P, Hatas E, Ghurye J, Maguire J, Mahmoud M, Cheng H, Heller D, Zook J M, Moemke T, Marschall T, Sedlazeck F J, Aach J, Chin C-S, Church G M, Li H. 2020. Chromosome-scale, haplotype-resolved assembly of human genomes. *Nature Biotechnology*, **39**, 309.

Garg V, Dudchenko O, Wang J, Khan A W, Gupta S, Kaur P, Han K, Saxena R K, Kale S M, Pham M, Yu J, Chitikineni A, Zhang Z, Fan G, Lui C, Valluri V, Meng F, Bhandari A, Liu X, Yang T, Chen H, Valliyodan B, Roorkiwal M, Shi C, Yang H B, Durand N C, Pandey M K, Li G, Barmukh R, Wang X, Chen X, Lam H-M, Jiang H, Zong X, Liang X, Liu X, Liao B, Guo B, Jackson S, Nguyen H T, Zhuang W, Shubo W, Wang X, Aiden E L, Bennetzen J L, Varshney R K. 2022. Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement. *Journal of Advanced Research*, **42**, 315-329.

Gao Y, Yang X, Chen H, Tan X, Yang Z, Deng L, Wang B, Kong S, Li S, Cui Y, Lei C, Wang Y, Pan Y, Ma S, Sun H, Zhao X, Shi Y, Yang Z, Wu D, Wu S, Zhao X, Shi B, Jin L, Hu Z; Chinese Pangenome Consortium (CPC); Lu Y, Chu J, Ye K, Xu S. 2023. A pangenome reference of 36 Chinese populations, *Nature*, **619**,112-121.

Garg V, Bohra A, Mascher M, Spannagl M, Xu X, Bevan M W, Bennetzen J L, Varshney R K. 2024. Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nature Genetics*, **56**, 1788-1799.

Gershman A, Sauria M E G, Guitart X, Vollger M R, Hook P W, Hoyt S J, Jain M, Shumate A, Razaghi R, Koren S, Altemose N, Caldas G V, Logsdon G A, Rhie A, Eichler E E, Schatz M C, O'Neill R J, Phillippy A M, Miga K H, Timp W. 2022. Epigenetic patterns in a complete human

genome. *Science*, **376**, eabj5089.

Guan D, McCarthy S A, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, **36**,2896-2898

Guha P, Das A, Dutta S, Chaudhuri T K. 2018. A rapid and efficient DNA extraction protocol from fresh and frozen human blood samples. *Journal of Clinical Laboratory Analysis*, **32**, e22181

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072-1075.

Haas B J, Salzberg S L, Zhu W, Pertea M, Allen J E, Orvis J, White O, Buell C R, Wortman J R. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, **9**, R7.

Huang Z, Xu Z, Bai H, Huang Y, Kang N, Ding X, Liu J, Luo H, Yang C, Chen W, Guo Q, Xue L, Zhang X, Xu L, Chen M, Fu H, Chen Y, Yue Z, Fukagawa T, Liu S, Chang G, Xu L. 2023. Evolutionary analysis of a complete chicken genome. *Proceedings of the National Academy of Sciences of the United States of America*, **120**, e2216641120.

Igolkina A A, Vorbrugg S, Rabanal F A, Liu H J, Ashkenazy H, Kornienko A E, Fitz J, Collenberg M, Kubica C, Mollá Morales A, Jaegle B, Wrightsman T, Voloshin V, Bezlepsky A D, Llaca V, Nizhynska V, Reichardt I, Bezrukov I, Lanz C, Bemm F, Flood P J, Nemomissa S, Hancock A, Guo Y L, Kersey P, Weigel D, Nordborg M. 2025. A comparison of 27 Arabidopsis thaliana genomes and the path toward an unbiased characterization of genetic polymorphism. *Nature Genetics*, **57**, 2289-2301.

Jevit M J, Castaneda C, Paria N, Das P J, Miller D, Antczak D F, Kalbfleisch T S, Davis B W, Raudsepp T. 2023. Trio-binning of a hinny refines the comparative organization of the horse and donkey X chromosomes and reveals novel species-specific features. *Scientific Reports*, **13**, 20180.

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, Chuong E, Clark E, Clarke S, Cockett N, Couldrey C, Davis B W, Elsik C G, Faraut T, Gao Y, Genet C, Grady P, Green J, Green R, Guan D, Hagen D, Hartley G A, Heaton M, Hoyt S J, Huang W, Jarvis E, Kalleberg J, Khatib H, Koepfi K-P, Koltes J, Koren S, Kuehn C, Leeb T, Leonard A, Liu G E, Low W Y, McConnell H, McRae K, Miga K, Mousel M, Neibergs H, Olagunju T, Pennell M, Petry B, Pewsner M, Phillippy A M, Pickett B D, Pineda P, Potapova T, Rachagani S, Rhie A, Rijnkels M, Robic A, Rodriguez Osorio N, Safonova Y, Schettini G, Schnabel R D, Natesh N S, Stegemiller M, Storer J, Stothard P, Stull C, Tosser-Klopp G, Traglia G M, Tuggle C K, Van Tassell C P, Watson C, Weikard R, Wimmers K, Xie S, Yang L, Smith T P L, O'Neill R J, Rosen B D. 2024. The Ruminant Telomere-to-Telomere (RT2T) Consortium. *Nature Genetics*, **56**, 1566-1573.

Koren S, Rhie A, Walenz BP, Dilthey A T, Bickhart D M, Kingan S B, Hiendleder S, Williams J L, Smith T P L, Phillippy A M. 2018. De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology*, **22**:10.1038/nbt.4277.

Lalli J L, Bortvin A N, McCoy R C, Werling D M. 2025. A T2T-CHM13 recombination map and globally diverse haplotype reference panel improves phasing and imputation. *bioRxiv* [Preprint], 639687.

Lam E T, Hastie A, Lin C, Ehrlich D, Das SK, Austin M D, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok P Y. 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nature Biotechnology*, **30**, 771-776.

Li H, Bloom J M, Farjoun Y, Fleharty M, Gauthier L, Neale B, MacArthur D. 2018. A synthetic-diploid benchmark for accurate variant-calling evaluation. *Nature Methods*, **15**, 595-597.

Li H, Durbin R. 2024. Genome assembly in the telomere-to-telomere era. *Nature Reviews Genetics*, **25**, 658–670.

Li M, Sun C, Xu N, Bian P, Tian X, Wang X, Wang Y, Jia X, Heller R, Wang M, Wang F, Dai X, Luo R, Guo Y, Wang X, Yang P, Hu D, Liu Z, Fu W, Zhang S, Li X, Wen C, Lan F, Siddiki A Z , Suwannapoom C, Zhao X, Nie Q Hu X, Jiang Y, Yang N. 2022. De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. *Molecular Biology and Evolution*, **39**, msac066.

Li R, Gong M, Zhang X, Wang F, Liu Z, Zhang L, Yang Q, Xu Y, Xu M, Zhang H, Zhang Y, Dai X, Gao Y, Zhang Z, Fang W, Yang Y, Fu W, Cao C, Yang P, Ghanatsaman Z A, Negari N J, Nanaei H A, Yue X, Song Y, Lan X, Deng W, Wang X, Pan C, Xiang R, Ibeagha-Awemu E M, Heslop-Harrison P J S, Rosen B D, Lenstra J A, Gan S, Jiang Y. 2023. A sheep pangenome reveals the spectrum of structural variations and their effects on tail phenotypes. *Genome research*, **33**, 463-477.

Lieberman-Aiden E, van Berkum N L, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie B R, Sabo P J, Dorschner M O, Sandstrom R, Bernstein B, Bender M A, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny L A, Lander E S, Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289-293.

Lin Y, Ye C, Li X, Chen Q, Wu Y, Zhang F, Pan R, Zhang S, Chen S, Wang X, Cao S, Wang Y, Yue Y, Liu Y, Yue J. 2023. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research*, **10**, uhad127.

Liu X, Liu W, Lenstra J A, Zheng Z, Wu X, Yang J, Li B, Yang Y, Qiu Q, Liu H, Li K, Liang C, Guo X, Ma X, Abbott R J, Kang M, Yan P, Liu J. 2023. Evolutionary origin of genomic structural variations in domestic yaks. *Nature Communication*, **14**, 5617.

Logsdon G A, Ebert P, Audano P A, Loftus M, Porubsky D, Ebler J, Yilmaz F, Hallast P, Prodanov T, Yoo D, Paisie C A, Harvey W T, Zhao X, Martino G V, Henglin M,

Munson K M, Rabbani K, Chin C S, Gu B, Ashraf H, Scholz S, Austine-Orimoloye O, Balachandran P, Bonder M J, Cheng H, Chong Z, Crabtree J, Gerstein M, Guethlein L A, Hasenfeld P, Hickey G, Hoekzema K, Hunt S E, Jensen M, Jiang Y, Koren S, Kwon Y, Li C, Li H, Li J, Norman P J, Oshima K K, Paten B, Phillippy A M, Pollock N R, Rausch T, Rautiainen M, Song Y, Söylev A, Sulovari A, Surapaneni L, Tsapalou V, Zhou W, Zhou Y, Zhu Q, Zody M C, Mills R E, Devine S E, Shi X, Talkowski M E, Chaisson M J P, Dilthey A T, Konkel M K, Korbel J O, Lee C, Beck C R, Eichler E E, Marschall T. 2025. Complex genetic variation in nearly complete human genomes. *Nature*, **644**, 430-441.

Logsdon G A, Rozanski A N, Ryabov F, Potapova T, Shepelev V A, Catacchio C R, Porubsky D, Mao Y, Yoo D, Rautiainen M, Koren S, Nurk S, Lucas J K, Hoekzema K, Munson K M, Gerton J L, Phillippy A M, Ventura M, Alexandrov I A, Eichler E E. 2024. The variation and evolution of complete human centromeres. *Nature*, **629**, 136-145.

Lu J, Huang P, Sun J, Liu J. 2023. DupScan: predicting and visualizing vertebrate genome duplication database. *Nucleic Acids Research*, **51**, D906-D912.

Luo L Y, Wu H, Zhao L-M, Zhang Y H, Huang J H, Liu Q Y, Wang H T, Mo D X, EEr H H, Zhang L Q, Chen H L, Jia S-G, Wang W-M, Li M-H. 2025. Telomere-to-telomere sheep genome assembly identifies variants associated with wool fineness. *Nature Genetics*, **57**, 218-230.

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, Sun R P, Wang F, Jia S G, Fang M Y. 2026. Telomere-to-telomere genome assembly of a male pig provides insight into population structure and selection for body stature. *Nature Genetics*, **58**, 195-205.

Manni M, Berkeley M R, Seppey M, Simão F A, Zdobnov E M. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, **38**, 4647-4654.

Mc Cartney A M, Shafin K, Alonge M, Bzikadze A V, Formenti G, Fungtammasan A, Howe K, Jain C, Koren S, Logsdon G A, Miga K H, Mikheenko A, Paten B, Shumate A, Soto D C, Sović I, Wood J M D, Zook J M, Phillippy A M, Rhie A. 2022. Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies. *Nature Methods*, **19**, 687-695.

Meng Q, Ochoa I, Hernaez M. 2020. GPress: a framework for querying general feature format (GFF) files and expression files in a compressed form. *Bioinformatics*, **36**, 4810-4812.

Miga, K. H， Sullivan, B. A. 2021. Expanding studies of chromosome structure and function in the era of T2T genomics. *Human Molecular Genetics*, **30**, R198–R205.

Miller D E, Staber C, Zeitlinger J, Hawley R S. 2018. Highly Contiguous Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing. *G3 (Bethesda)*, **8**, 3131-314.

Nguinkal J A, Zoclanclounon Y A B, Brunner R M, Chen Y, Goldammer T. 2024. Haplotype-resolved and near-T2T genome assembly of the African catfish (Clarias gariepinus). *Scientific Data*, **11**, 109.

Nguyen T V, Vander Jagt C J, Wang J, Daetwyler H D, Xiang R, Goddard M E, Nguyen L T, Ross E M, Hayes B J, Chamberlain A J, MacLeod I M. 2023. In it for the long run: perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants. *Genetics Selection Evolution*, **55**, 9.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V , Mikheenko A, Vollger M R, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt S J, Diekhans M, Logsdon G A, Alonge M, Antonarakis S E, Borchers M, Bouffard G G, Brooks S Y, Caldas G V, Chen N-C, Cheng H, Chin C-S, Chow W, de Lima L G , Dishuck P C, Durbin R, Dvorkina T, Fiddes I T, Formenti G, Fulton R S, Fungtammasan A, Garrison E, Grady P G S, Graves-Lindsay T A , Hall I M, Hansen N F, Hartley G A, Haukness M, Howe K, Hunkapiller M W9, Jain C, Jain M, Jarvis E D, Kerpedjiev P, Kirsche M, Kolmogorov M, Korlach J, Kremitzki M, Li H, Maduro V V, Marschall T, McCartney A M, McDaniel J, Miller D E, Mullikin J C, Myers E W, Olson N D, Paten B, Peluso P, Pevzner P A, Porubsky D, Potapova T, Rogaev E I, Rosenfeld J A, Salzberg S L, Schneider V A, Sedlazeck F J , Shafin K, Shew C J, Shumate A, Sims Y, Smit A F A , Soto D C, Sović I , Storer J M , Streets A, Sullivan B A, Thibaud-Nissen F, Torrance J, Wagner J, Walenz B P, Wenger A, Wood J M D, Xiao C, Yan S M, Young A C, Zarate S, Surti U , McCoy R C, Dennis M Y, Alexandrov I A, Gerton J L, O'Neill R J, Timp W, Zook J M, Schatz M C, Eichler E E, Miga K H, Phillippy A M. 2022. The complete sequence of a human genome. *Science*, **376**, 44-53.

Olagunju T A, Rosen B D, Neibergs H L, Becker G M, Davenport K M, Elsik C G, Hadfield T S, Koren S, Kuhn K L, Rhie A, Shira K A, Skibiel A L, Stegemiller M R, Thorne J W, Villamediana P, Cockett N E, Murdoch B M, Smith T P L. 2024. Telomere-to-telomere assemblies of cattle and sheep Y-chromosomes uncover divergent structure and gene content. *Nature Communication*, **15**, 8277.

Oppenheimer J, Rosen B D, Heaton M P, Vander Ley B L, Shafer W R, Schuetze F T, Stroud B, Kuehn L A, McClure J C, Barfield J P, Blackburn H D, Kalbfleisch T S, Bickhart D M, Davenport K M, Kuhn K L, Green R E, Shapiro B, Smith T P L. 2021. A Reference Genome Assembly of American Bison, Bison bison bison. *Journal of Heredity*, **112**, 174–183.

Pedersen B S, Quinlan A R. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, **34**, 867- 868.

Pineda P S, MacPhillamy , Ren Y, Chen T, Zhong L, Adelson D L, Dessaix C, Perez-Silva J, Haggerty L, Martin F J, Bottema C D K, Pitchford W S, Rosen B D, Smith T P L, Low W Y. 2025. Insights into natural neocentromere

evolution from a cattle T2T X chromosome. *Nature Communication*, **16**, 10745.

Piro V C, Faoro H, Weiss V A, Steffens M B R, Pedrosa F O, Souza E M, Raittz R T. 2014. FGAP: an automated gap closing tool. *BMC Research Notes*, **7**, 371.

Porto-Foresti F, Oliveira C, Tabata Y A, Rigolino M G, Foresti F. 2002. NORs inheritance analysis in crossings including individuals from two stocks of rainbow trout (Oncorhynchus mykiss). *Hereditas*, **136**, 227-230.

Quinlan A R, Hall I M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841- 842.

Rautiainen M, Nurk S, Walenz B P, Logsdon G A, Porubsky D, Rhie A, Eichler E E, Phillippy A M, Koren S. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nature Biotechnology*, **41**, 1474-1482.

Rhie A, Nurk S, Cechova M, Hoyt S J, Taylor D J, Altemose N, Hook P W, Koren S, Rautiainen M, Alexandrov I A, Allen J, Asri M, Bzikadze A V, Chen N C, Chin C S, Diekhans M, Flicek P, Formenti G, Fungtammasan A, Garcia Giron C, Garrison E, Gershman A, Gerton J L, Grady P G S, Guarracino A, Haggerty L, Halabian R, Hansen NF, Harris R, Hartley G A, Harvey W T, Haukness M, Heinz J, Hourlier T, Hubley R M, Hunt S E, Hwang S, Jain M, Kesharwani R K, Lewis A P, Li H, Logsdon G A, Lucas J K, Makalowski W, Markovic C, Martin F J, Mc Cartney A M, McCoy R C, McDaniel J, McNulty B M, Medvedev P, Mikheenko A, Munson K M, Murphy T D, Olsen H E, Olson N D, Paulin L F, Porubsky D, Potapova T, Ryabov F, Salzberg S L, Sauria M E G, Sedlazeck F J, Shafin K, Shepelev V A, Shumate A, Storer JM, Surapaneni L, Taravella Oill AM, Thibaud-Nissen F, Timp W, Tomaszkiewicz M, Vollger M R, Walenz B P, Watwood A C, Weissensteiner M H, Wenger A M, Wilson M A, Zarate S, Zhu Y, Zook J M, Eichler E E, O'Neill R J, Schatz M C, Miga K H, Makova K D, Phillippy A M. 2023. The complete sequence of a human Y chromosome. *Nature*, **621**, 344-354.

Rhie A, Walenz B P, Koren S, Phillippy A M. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, **21**, 245.

Roach M J, Schmidt S A, Borneman A R. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, **19**, 460.

Schwartz D C, Cantor C R. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, **37**, 67–75.

Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, Nater A, Bureš S, Garamszegi L Z, Hogner S, Moreno J, Qvarnström A, Ružić M, Sæther S A, Sætre G P, Török J, Ellegren H. 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nature Communication*, **6**, 7330.

Smith T, Olagunju T, Rosen B, Neibergs H, Becker G, Davenport K, Elsik C, Hadfield T, Koren S, Kuhn K, Rhie A, Shira K, Skibiel A, Stegemiller M, Thorne J, Villamediana P, Cockett N, Murdoch B. 2024. The first complete T2T Assemblies of Cattle and Sheep Y-Chromosomes uncover remarkable divergence in structure and gene content. *Research Square* [Preprint], rs.3.rs-4033388.

Su R, Zhou H, Yang W, Moqir S, Ritu X, Liu L, Shi Y, Dong A, Bayier M, Letu Y, Manxi X, Chulu H, Nasenochir N, Meng H, Herrid M. 2024. Near telomere-to-telomere genome assembly of Mongolian cattle: implications for population genetic variation and beef quality. *GigaScience*, **13**, giae099.

Tsukahara S, Bousios A, Perez-Roman E, Yamaguchi S, Leduque B, Nakano A, Naish M, Osakabe A, Toyoda A, Ito H, Edera A, Tominaga S, Juliarni, Kato K, Oda S, Inagaki S, Lorković Z, Nagaki K, Berger F, Kawabe A, Quadrana L, Henderson I, Kakutani T. 2025. Centrophilic retrotransposon integration via CENH3 chromatin in Arabidopsis. *Nature*, **637**, 744-748.

Turner I, Garimella K V, Iqbal Z, McVean G. 2018. Integrating long-range connectivity information into de Bruijn graphs. *Bioinformatics*, **34**, 2556-256.

Varshney R K, Sinha P, Singh V K, Kumar A, Zhang Q, Bennetzen J L. 2020. 5Gs for crop genetic improvement. *Current Opinion in Plant Biology*, **56**, 190-196.

Vollger M R, Dishuck P C, Harvey W T, DeWitt W S, Guitart X, Goldberg M E, Rozanski A N, Lucas J, Asri M; Human Pangenome Reference Consortium; Munson K M, Lewis A P, Hoekzema K, Logsdon G A, Porubsky D, Paten B, Harris K, Hsieh P, Eichler E E. 2023. Increased mutation and gene conversion within human segmental duplications. *Nature*, **617**, 325-334.

Vurture G W, Sedlazeck F J, Nattestad M, Underwood C J, Fang H, Gurtowski J, Schatz M C. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**, 2202–2204.

Wang B, Yang X, Jia Y, Xu Y, Jia P, Dang N, Wang S, Xu T, Zhao X, Gao S, Dong Q, Ye K. 2022. High-quality Arabidopsis thaliana Genome Assembly with Nanopore and HiFi Long Reads. *Genomics Proteomics Bioinformatics*, **20**, 4-13.

Wang Y, Zhao Y, Bollas A, Wang Y, Au K F. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, **39**,1348-1365.

Wenger A M, Peluso P, Rowell W J, Chang P-C, Hall R J, Concepcion G T, Ebler J, Fungtammasan A, Kolesnikov A, Olson N D, Töpfer A, Alonge M, Mahmoud M, Qian Y, Chin C S, Phillippy A M, Schatz M C8, Myers G, DePristo M A, Ruan J, Marschall T, Sedlazeck F J, Zook J M, Li H, Koren S, Carroll A, Rank D R, Hunkapiller M W. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, **37**, 1155-1162.

Wen H, Yang J, Zhao X, Wang X, Lei J, Li Y, Du W, Li D, Xu Y, Lonardi S, Pan W. 2025. TRFill: synergistic use of HiFi and Hi-C sequencing enables accurate assembly of tandem repeats for population-level analysis. *Genome Biology*, **26**, 227.

Wu H, Luo LY, Zhang Y H, Zhang C Y, Huang J H, Mo D X, Zhao L M, Wang Z X, Wang Y C, He-Hua E, Bai W L, Han D, Dou X T, Ren Y L, Dingkao R, Chen H L, Ye Y, Du H D, Zhao Z Q, Wang X J, Jia S G, Liu Z H, Li M H. 2024. Telomere-to-telomere genome assembly of a male goat reveals variants associated with cashmere traits. *Nature Communication*, **15**, 10041.

Wuyun B, Liu J, Wu R, Li Y, Liu F, Zhao H, Hao C, Zhao G, Sun W, Song Y, Wang W, Wang Y, Ma C, Xu F, He J, Wang P, Bao X, Cao G, Zhang Y, Lu Y, Li X. 2026. A telomere-to-telomere gapless genome assembly of the Tibetan wild ass (Equus kiang). *Scientific Data*, **13**:182.

Xu M, Guo L, Gu S, Wang O, Zhang R, Peters B A, Fan G, Liu X, Xu X, Deng L, 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.

You X, Fang Q, Chen C, Cao J, Fu S, Zhang T, Wang S, He X, He J, Zhou Y, Wang B, Wang L, Wang Z, Sun T,Yang X, Te R, Jian J, Zhou H, Dai Y, Liu Y. 2024. A sheep genome. *Scientific Data*, **11**, 762.

Zhang X, Chen J, Zhou W, Wen J, Shi Q. 2025. A telomere-to-telomere gap-free genome assembly of the protandrous hermaphrodite Asian seabass (Lates calcarifer). *Scientific Data*, **12**, 1457.

Zhao H, Zhou H, Sun G, Dong B, Zhu W, Mu X, Li X, Wang J, Zhao M, Yang W, Zhang G, Ji R, Geng T, Gong D, Meng H, Wang J. 2024. Telomere-to-telomere genome assembly of the goose Anser cygnoides. *Scientific Data*, **11**, 741.

Zhao Q, Yin Z, Hou Z. 2025. Near telomere-to-telomere genome assemblies of Silkie Gallus gallus and Mallard Anas platyrhynchos restored the structure of chromosomes and “missing” genes in birds. *Journal of Animal Science and Biotechnology*, **16**, 9.

Zhou Z W, Yu Z G, Huang X M, Liu J S, Guo Y X, Chen L L, Song J M. 2022. GenomeSyn: a bioinformatics tool for visualizing genome synteny and structural variations. *Journal of Genetics and Genomics*, **49**, 1174–1176.

Zong W, Chen L, Zhang D, Zhang Y, Wang J, Hou X, Chai J, An Y, Tian M, He X, Song C, He J, Liu X, Wang L, D'Alessandro E, Wang L, Yin Y, Li M, Liu D, Wang J, Zhang L. 2025. Two telomere-to-telomere pig genome assemblies and pan-genome analyses provide insights into genomic structural landscape and genetic adaptations. *Imeta*, **4**, e70013.

Executive Editor-in-Chief Xueyong Zhang
Managing Editor Ning Wang