Agri Articles
(e-Magazine for Agricultural Articles)
Volume: 05, Issue: 05 (SEP-OCT, 2025)
Available online at http://www.agriarticles.com
©Agri Articles, ISSN: 2582-9882

Long-Read Sequencing: Applications in Animal Genetics

*Kankan Baro[1], Ranjith G[1], Yogashmita Boro[2] and Manish Borah[3]
[1]Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Bareilly, UP
[2]Division of Livestock Products Technology, ICAR- Indian Veterinary Research Institute, Bareilly, UP, India
[3]Department of Livestock Production & Management, CVSc, Khanapara, AAU, Ghy-22, Assam, India
*Corresponding Author's email: drkankanbaro@gmail.com

Since the advent of sequencing technologies, continuous advancements have significantly enhanced their accuracy, speed, and cost-efficiency, establishing them as indispensable tools in genetic research. The long-read sequencing technologies, capable of generating reads exceeding 10 kilobases, have revolutionized genomic research in livestock species. These technologies have provided deeper insights into genome organization, structural variation, functional elements, and evolutionary dynamics. Compared to traditional short-read sequencing platforms, which produce significantly shorter fragments, long-read sequencing platforms offer superior resolution for complex genomic regions, thereby enhancing the accuracy of genome assembly and annotation and has proved to become an essential tool for animal genetics studies.

Keywords: Long-read sequencing, PacBio, ONT, DNA sequencing.

Introduction

The credit for the invention of DNA sequencing, the process of determining the order of nucleotide bases of a DNA molecule or genome, is attributed to Frederick Sanger and colleagues, who first demonstrated the dideoxy chain-termination (Sanger) method in 1977. In the same year, Albert Maxam and Walter Gilbert introduced a DNA sequencing technique based on chemical degradation, which became widely known as Maxam–Gilbert sequencing. Those two sequencing technologies are considered as first generation sequencing. Next-generation sequencing (NGS) was developed to overcome the limitations of first-generation sequencing technologies. Its affordability and capacity to generate large volumes of data have led to its widespread adoption as a standard tool in genomics. Next-generation sequencing (NGS) technologies have significantly advanced genomic research by enabling highly accurate sequencing. Among the NGS technologies, Second-generation sequencing, also known as short-read sequencing (SRS), typically produces reads ranging from 100 to 300 base pairs (bp). While effective for many applications, the short read length can pose challenges in assembling complex genomes and detecting structural variants. To address these limitations, third-generation sequencing technologies, capable of generating reads exceeding 10 kilobases (kb), commonly referred to as long-read sequencing (LRS)—have been developed. These platforms offer improved genome assembly, structural variant detection, and superior resolution of repetitive regions.

Long-read sequencing

The sequencing process by which >10kb reads are generated is called long-read sequencing and the reads are called as "long reads". Long-read sequencing is also known as third generation sequencing (TGS). In this technology, sequencing of single DNA molecule is

done without the need to halt between read steps (whether enzymatic or otherwise). Two of the currently famous long-read sequencing technologies are SMRT sequencing and ONT. SMRT, or single-molecule real-time sequencing, also referred to as PacBio sequencing due to its development by Pacific Biosciences, works by detecting fluorescent signals generated during the incorporation of individual nucleotides by a polymerase enzyme fixed at the bottom of a tiny well. ONT, or Oxford Nanopore Technologies, uses nanopore sequencers that detect variations in ionic current as single-stranded nucleic acids pass through biological nanopores. Since each nucleotide produces a unique resistance level, the resulting current patterns can be analysed to determine the sequence of bases. In LRS it is possible to sequence reads that are orders of magnitude longer than the short reads sequenced in SRS. SMRT can sequence 10-20 kb highly accurate reads and lower-accuracy reads up to hundreds of kilobase pairs. While ONT can sequence even longer reads of up to 1–2 mb.

Advantages of long-read sequencing

Short read lengths make it difficult to sequence certain repetitive regions in the genome, which can cause errors in assembly and leave gaps in the final genome sequence. Moreover, its inability to directly sequence DNA or RNA without the clonal amplification hinders its capacity to meet the growing technological requirements of contemporary biological research. While in long reads sequencing it is possible to generate long reads (>10kbp) and sequence the DNA or RNA without a prior amplification step. Also, real-time base sequencing can be done and shorter processing time is needed. It efficiently performs comprehensive sequencing of the livestock genome, generating a complete genomic dataset that supports the identification of genetic variations and the understanding of gene functions. Long-read sequencing provides greater accuracy in detecting structural variations in the genome—such as insertions, deletions, and inversions—delivering more precise genomic data for studying important traits. Additionally, in epigenetic research, the generated long-read allow for a more detailed and accurate analysis of DNA methylation and RNA modification patterns. Furthermore, when applied to genomic selection, this technology enables the thorough and accurate examination of large-scale individual genome data, supporting selection strategies that promote faster genetic progress and improve breeding efficiency in livestock.

Applications in Animal Genetics

Long-read sequencing has become a valuable tool for uncovering genome structure, function, and evolutionary patterns. This technology now enables the detailed analysis of structural variants like never before and supports the development of methods to assess their importance in key livestock traits. Chang et al. used PacBio SMRT sequencing to analyse the full-length transcriptome of cattle. This approach enabled the identification of numerous alternative splicing events, alternative polyadenylation sites, novel isoforms, previously unannotated long non-coding RNAs, and transcription factors, offering a more complete understanding of transcriptome diversity in cattle. By using long-read sequencing Fang et al. identified numerous genes, including PLK2, CNN1, TNFSF15, POSTN, TNIP1, N4BP2L1, and PRKAA1, as key candidates involved in bovine adipocyte differentiation. These findings offer new insights into fat accumulation and contribute to enhancing the meat quality of Chinese red steppe cattle. The Oxford Nanopore Technologies' minION sequencer was utilized by Lamb et al. to sequence four homozygous poll (PcPc), four homozygous horned (pp), and three heterozygous (Pcp) Brahman cattle to analyse the poll allele in this breed. Analysis of the mapped reads at the poll locus showed insertions of about 200 base pairs in the poll animals that were not present in the horned animals. In a separate study by Mukherjee et al., a hybrid approach combining PacBio long-reads and Illumina paired-end reads was used to assemble the Indian Mithun genome. They reported that this assembly offers improved coverage, is less fragmented, better annotated, and provides a more complete genome compared to the previously published Gayal genome.

Challenges and Limitations

Long-read sequencing platforms have a high error rate and relatively low throughput. Therefore, enhancing the performance of these two platforms (PacBio & ONT) requires future models to focus on reducing error rates and boosting data output. To further address and correct the relatively high error rates commonly associated with PacBio long-read sequencing, Chang et al. supplemented their dataset by generating high-accuracy Illumina short reads. This combination of technologies allowed them to retain the advantages of long-read sequencing while significantly improving overall sequence accuracy through short-read correction. Long-read sequencing (LRS) throughput is greatly affected by molecular damage during library preparation, making it essential to use high-quality, intact, high-molecular-weight DNA. Precise size selection of this DNA is crucial, as LRS platforms tend to favour smaller, faster-diffusing molecules during sequencing, which can introduce loading bias and result in inconsistent read lengths.

Conclusion

In the past few years long-read sequencing technologies have been widely used to sequence DNA of many organisms' including livestock species and will continue to be used in future with better accuracy and cheaper cost. This new sequencing technologies could take greater advantage of DNA polymerase's high catalytic efficiency and processivity, or even eliminate the need for biological or chemical processes entirely, leading to significantly longer read lengths and much faster results—potentially reducing sequencing time from days to just hours or minutes and thus expanding the use of these technologies across nearly all areas of life and biomedical sciences.

References

1. Chang, T., An, B., Liang, M., Duan, X., Du, L., Cai, W., ... & Li, J. (2021). PacBio single-molecule long-read sequencing provides new light on the complexity of full-length transcripts in cattle. Frontiers in Genetics, 12, 664974
2. Fang, W., Yang, S., Li, X., Nanaei, A., Liu, Y., Cao, Y., ... & Cao, Y. (2025). Long-read sequencing uncovers key regulatory genes involved in the differentiation of preadipocytes of Chinese red steppe cattle. Scientific Reports, 15(1), 29459.
3. Lamb, H. J., Ross, E. M., Nguyen, L. T., Lyons, R. E., Moore, S. S., & Hayes, B. J. (2020). Characterization of the poll allele in Brahman cattle using long-read Oxford Nanopore sequencing. Journal of animal science, 98(5), skaa127.
4. Liu, X., Zheng, J., Ding, J., Wu, J., Zuo, F., & Zhang, G. (2024). When livestock genomes meet third-generation sequencing technology: From opportunities to applications. Genes, 15(2), 245.
5. Mukherjee, S., Cai, Z., Mukherjee, A., Longkumer , I., Mech, M., Vupru, K., ... & Sahana, G. (2019). Whole genome sequence and de novo assembly revealed genomic architecture of Indian Mithun (Bos frontalis). BMC genomics, 20(1), 1-12.