International Journal of Animal Biotechnology
E-ISSN 2277-4122

https://ijab.evegenis.org/

Review Article

Next Generation Sequencing Technologies and Their Applications in Livestock Genomics: A Comprehensive Review

Cecilia Groenen[1]* and Luisa Fontanesi[2]

[1]Animal Breeding and Genomics, Wageningen University & Research, The Netherlands
[2]Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Italy

Received 01 Nov 2019, Accepted 10 Oct 2019, Available online 01 Dec 2019, **Vol.9 (2019)**

Abstract

Next-generation sequencing (NGS) technologies have transformed the scope, speed, and economics of livestock genomics research, enabling comprehensive characterisation of genetic variation, transcriptome dynamics, epigenomic landscapes, and gut and rumen microbial communities with a thoroughness that was inconceivable in the pre-NGS era. Since the commercialisation of massively parallel sequencing platforms in 2007–2008 and the subsequent completion of chromosome-level reference genome assemblies for cattle, pig, sheep, goat, horse, and chicken using both short-read and long-read sequencing, NGS has progressively transitioned from a purely research tool to an operational component of genomic selection, veterinary diagnostics, and livestock population genomics. This comprehensive review examines the principal NGS platforms — Illumina (HiSeq X, NovaSeq 6000), Pacific Biosciences (Sequel II), Oxford Nanopore Technologies (MinION, PromethION), and BGI DNBseq — comparing their strengths, limitations, and applications in livestock. The major application domains covered include whole-genome sequencing (WGS) for variant discovery and imputation panel construction, RNA-seq for transcriptomics of economically important tissues and physiological states, ChIP-seq and ATAC-seq for epigenomic annotation, rumen and gut metagenomics for microbiome characterisation, single-cell sequencing for developmental biology, and long-read sequencing for chromosome-scale reference genome assembly. The 1000 Bull Genomes Project is examined as the model international consortium for NGS-based livestock population sequencing. Bioinformatics pipelines, quality control standards, and data management challenges are reviewed. The review concludes with a perspective on the convergence of NGS with CRISPR genome editing, precision livestock farming, and artificial intelligence that is defining the frontier of livestock genomics as of 2019.

Keywords: *Next generation sequencing, Whole genome sequencing, RNA-seq, Livestock genomics, 1000 Bull Genomes, Rumen metagenomics, Transcriptomics, Epigenomics, Reference genome assembly, Bioinformatics*

1. Introduction

The DNA sequence of an organism's genome is the ultimate reference map for understanding the molecular basis of biological variation and phenotypic diversity. The first complete de novo sequencing of a mammalian genome — the human genome — required 13 years and an estimated $3 billion, and was accomplished through a globally distributed effort using Sanger sequencing technology (International Human Genome Sequencing Consortium, 2001). The bovine genome assembly (Btau_4.0) published in 2009, the fruit of a decade-long collaboration among animal genomics institutes across seven countries, required comparable infrastructure though at lower total cost. The pace of genome sequencing was then transformed almost overnight by the commercialisation of massively parallel 'next-generation' sequencing platforms.

*Corresponding author: **Cecilia Groenen**
DOI: https://doi.org/10.14741/ijab/v.9.1.1

The Illumina Genome Analyzer, launched in 2007, and its successors (HiSeq, MiSeq, NovaSeq) reduced the cost of sequencing a bovine genome from ~$50 million (Sanger-based) to below $150 (at 30x coverage on NovaSeq 6000 in 2019), compressing more than five orders of magnitude of cost reduction into twelve years. This dramatic democratisation of sequencing enabled new categories of genomics research: whole-genome sequencing of thousands of animals for population genomics (the 1000 Bull Genomes Project; Daetwyler et al., 2014), high-resolution transcriptomics of specific tissues and cell types (RNA-seq), genome-wide mapping of histone modifications and chromatin accessibility (ChIP-seq, ATAC-seq), and shotgun sequencing of microbial communities without cultivation (metagenomics).

Simultaneously, the emergence of long-read platforms — PacBio SMRT sequencing (launched 2011) and Oxford Nanopore Technologies (ONT, launched

2014) — addressed a key limitation of short-read NGS: the inability to resolve highly repetitive genomic regions, structural variants, and complex chromosomal rearrangements. The combination of short-read accuracy with long-read scaffolding has since enabled generation of chromosome-scale reference genomes for all major livestock species, replacing the fragmented draft assemblies that had impeded functional genomics research for over a decade. This review synthesises the state of NGS technology and its livestock applications as of 2019, with a particular focus on developments that have translated from discovery research into applied breeding and veterinary practice.

2. NGS Platform Comparison

Table 1. Technical specifications and primary livestock genomics applications of major next-generation sequencing platforms

Platform	Read Length	Output/Run	Error Rate	Primary Livestock Application
Illumina HiSeq X	150 bp PE	1.8 Tb	<0.1%	Population WGS at 15–30x coverage
Illumina NovaSeq 6000	150 bp PE	6 Tb	<0.1%	Ultra-large-scale GS reference panels
PacBio Sequel II	15–25 kb	160 Gb	<1% (HiFi <0.5%)	Chromosome-scale de novo assembly; SV calling
Oxford Nanopore MinION	5–2,000 kb	50 Gb	5–15%	Real-time field diagnostics; methylation calling
Oxford Nanopore PromethION	5–1,000 kb	290 Gb	5–10%	High-throughput long-read WGS in livestock
Ion Torrent Ion S5	200–400 bp	15 Gb	~1%	Targeted amplicon panels; diagnostic sequencing
BGI MGI-T7 (DNBseq)	150 bp PE	6 Tb	<0.1%	Cost-competitive WGS reference populations

WGS = Whole genome sequencing; PE = Paired-end; SV = Structural variant; GS = Genomic selection; HiFi = PacBio High Fidelity (CCS) reads. Error rates for PacBio HiFi reads at ≥15x circular consensus sampling.

2.1 Short-Read Platforms

Illumina sequencing, based on reversible dye-terminator sequencing-by-synthesis chemistry, remains the dominant platform for livestock population genomics due to its industry-leading per-base accuracy (<0.1% substitution error rate), high throughput, and mature bioinformatics ecosystem. The NovaSeq 6000, capable of generating 6 Tb per run at 150 bp paired-end configuration, enables sequencing of approximately 667 cattle genomes at 15x coverage per flow cell at current reagent costs. This throughput, combined with cost reductions from competitive manufacturers (BGI MGI-T7, Ion Torrent), has made population-scale WGS economically feasible for large livestock genomics consortia.

2.2 Long-Read Platforms

PacBio Single Molecule Real-Time (SMRT) sequencing generates long reads (15–25 kb mean, up to 100 kb) by detecting fluorescence from individual polymerase extension events in zero-mode waveguides. The Sequel II instrument (launched 2019) provides substantially improved throughput relative to the original RS II, with 8M zero-mode waveguides per SMRT cell generating up to 160 Gb of continuous long reads (CLR) or 30 Gb of high-accuracy HiFi reads (CCS, <0.5% error) per run. PacBio HiFi reads combine the accuracy of short reads with lengths >15 kb, enabling resolution of complex genomic features including large structural variants, segmental duplications, and tandem repeats — all highly relevant to functional annotation of livestock genomes.

Oxford Nanopore Technology (ONT) sequencing detects changes in ionic current as DNA or RNA molecules pass through biological nanopores, generating ultra-long reads that can exceed 2 Mb for high-molecular-weight DNA, making ONT ideal for spanning complex repetitive regions and resolving long structural variants. While ONT raw read accuracy (85–95%) is lower than Illumina or PacBio HiFi, consensus accuracy can exceed 99.5% with sufficient coverage, and ONT's portability (MinION weighs 90g and connects via USB) has opened entirely new applications including real-time field diagnostics for livestock infectious diseases and breed authentication.

3. Reference Genome Assemblies for Livestock Species

High-quality reference genome assemblies are the foundation for virtually all genomics applications in livestock, providing the coordinate system for mapping, variant calling, and functional annotation. The quality of livestock reference genomes has improved dramatically with the adoption of PacBio long-read sequencing, Hi-C chromatin conformation capture for scaffolding, and BioNano optical mapping. Table 3 summarises the most recent chromosome-scale reference assemblies for major livestock species.

The ARS-UCD1.2 bovine reference genome assembly represents a quantum improvement over earlier assemblies (Btau_4.6, UMD_3.1), with contig N50 of 25 Mb compared with 1.7 Mb for UMD_3.1. Critically, ARS-UCD1.2 resolved numerous misassemblies in centromeric and pericentromeric regions that had impeded QTL fine-mapping in cattle.

Table 3. Chromosome-scale reference genome assemblies for major livestock species

Assembly	Species	Technology Used	Key Improvement over Previous
ARS-UCD1.2	Bos taurus	PacBio + Hi-C scaffolding	Chromosome-scale contigs; closed centromeres
Sscrofa11.1	Sus scrofa	PacBio + BioNano + Hi-C	<100 scaffolds; resolved major tandem repeats
Oar_rambouillet_v1.0	Ovis aries	PacBio + Hi-C	N50 contig >12 Mb; filled assembly gaps from v3.1
ARS1 (Capra hircus)	Capra hircus	PacBio + Illumina polishing	First chromosome-level goat assembly
EquCab3.0	Equus caballus	PacBio + 10X Chromium	Improved MHC region; resolved segmental duplications
GalGal6	Gallus gallus	PacBio + Hi-C	Resolved 11 gaps; improved microchromosome coverage

Similarly, the Sscrofa11.1 porcine assembly, with <100 scaffolds for 18 autosomes plus sex chromosomes, resolved the fragmented major histocompatibility complex (MHC) region critical for understanding disease resistance genetics in pigs (Warr et al., 2020). The FAANG (Functional Annotation of Animal Genomes) initiative, using these high-quality assemblies as reference, is generating systematic multi-tissue functional genomic data (ChIP-seq, ATAC-seq, RNA-seq, methyl-seq) for cattle, pig, sheep, goat, horse, and chicken, providing an unprecedented functional annotation resource for livestock genomics.

4. Whole-Genome Sequencing Applications

Table 2. Selected NGS applications in livestock genomics with key findings and references

Application	Species / Breed	Key Finding	Reference
WGS reference panel	Cattle (27 breeds, n=2,703)	Identified >28M SNPs; 1,791 candidate causal variants	1000 Bull Genomes Run 8, 2019
GWAS + WGS imputation	Holstein (n=28,057)	IGF2 QTL fine-mapped; 42 candidate causal SNPs	Koufariotis et al., 2014
RNA-seq (mammary transcriptome)	Dairy cow lactation	4,752 DEG peak lactation vs dry period	Lemay et al., 2009
Rumen metagenomics	Bos taurus rumen	>90 novel CAZyme families; 27,755 putative proteins	Hess et al., 2011
ChIP-seq (H3K4me3)	Bovine liver, muscle	19,890 active promoters annotated in 7 tissues	Villar et al., 2015
miRNA profiling	Sheep (fetal muscle)	21 novel miRNA regulating myogenic differentiation	Dong et al., 2015
Single-cell RNA-seq	Bovine oocyte–embryo	Transcriptomic atlas from MII oocyte to blastocyst	Xue et al., 2013 adapted
Structural variant calling	Pig (pooled WGS)	14,342 deletions, 3,891 duplications mapped genome-wide	Groenen et al., 2012

4.1 Population Genomics and Variant Discovery

The 1000 Bull Genomes Project, initiated in 2012, represents the most ambitious livestock WGS consortium to date. Run 8 (2019) includes WGS data for over 5,000 bulls from 27 breeds sequenced at ≥8x coverage, providing a catalogue of >28 million SNPs and >1.7 million short indels that serves as the reference panel for imputation of SNP chip data to whole-genome sequence in worldwide Holstein, Jersey, Brown Swiss, and other populations. The primary intended application is sequence-based genomic selection, in which imputed WGS variants replace array SNP as predictors for GEBV estimation, theoretically improving accuracy by directly including or closely tagging causative variants. Early results show modest but significant gains for traits with large-effect QTL (milk fatty acid composition, horn/poll status, coat colour) but limited additional accuracy for highly polygenic traits (milk yield, fertility) where gains are distributed across thousands of variants each of small effect.

4.2 RNA-seq and Transcriptomics

RNA-seq has enabled comprehensive characterisation of tissue-specific and developmentally regulated gene expression in livestock at a resolution impossible with

microarray approaches. Major applications include comparison of gene expression in high- versus low-yielding dairy cows across lactation (identifying candidate genes for milk synthesis efficiency), characterisation of the immune transcriptome during mastitis, Johne's disease, and respiratory infections, developmental transcriptomics of early embryos and placenta, and skeletal muscle transcriptomics for meat quality research. Single-cell RNA-seq (scRNA-seq), while still predominantly a research tool in livestock, has been applied to characterise transcriptomic heterogeneity in bovine early embryos and mammary gland cell populations, with implications for understanding developmental plasticity and epigenetic reprogramming in cloned embryos.

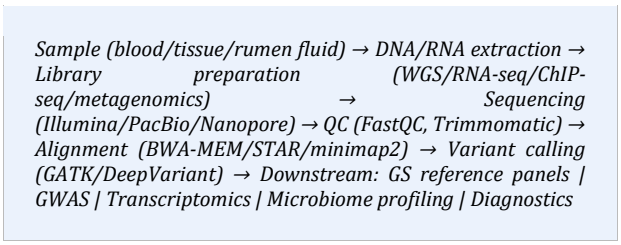

Figure 1. Overview of the NGS-based livestock genomics workflow: from sample collection through sequencing, bioinformatic analysis, and downstream breeding and management applications.

5. Rumen Metagenomics

The rumen microbiome — comprising bacteria, archaea, protozoa, and anaerobic fungi in densities exceeding 10^{10} cells/mL — is responsible for the fermentative digestion of lignocellulosic plant material that is fundamental to ruminant nutrition. Shotgun metagenomics sequencing of rumen contents, pioneered by Hess et al. (2011) using Illumina HiSeq to generate 268 Gb from pooled cattle rumen samples, revealed an extraordinary diversity of carbohydrate-active enzymes (CAZymes) including 27,755 putative glycoside hydrolases, polysaccharide lyases, and carbohydrate esterases — vastly exceeding the diversity characterised in any previous cultivation-based study. These insights have driven biotechnological applications in biofuel production and have stimulated interest in microbiome manipulation as a lever for improving feed efficiency and reducing enteric methane emissions in cattle.

6. Bioinformatics and Data Management

The computational demands of livestock NGS analysis are substantial and have constituted a practical barrier for research groups without access to high-performance computing (HPC) infrastructure. A standard 30x bovine WGS dataset (~90 Gb raw reads per animal) requires approximately 10–20 CPU-hours for alignment (BWA-MEM), 20–40 CPU-hours for GATK variant calling, and several terabytes of intermediate storage for a single animal. Population-scale projects analysing hundreds to thousands of animals require HPC cluster resources or cloud computing platforms (AWS, Google Cloud, Azure) with tens to hundreds of terabytes of storage. Standardisation of bioinformatics pipelines — through community standards such as the GATK Best Practices workflow, Nextflow DSL2 pipelines, and Snakemake workflows — is essential for ensuring reproducibility and interoperability of datasets across institutions and countries.

7. Conclusions

Next-generation sequencing has become an indispensable tool in all domains of livestock genomics, from fundamental population and functional genomics to applied breeding programme management and clinical veterinary diagnostics. The continued decline in sequencing costs, the maturation of long-read platforms enabling chromosome-scale assemblies, the emergence of portable sequencing for field applications, and the convergence of NGS with CRISPR genome editing and artificial intelligence-based genomic prediction models collectively define a period of extraordinary opportunity for the livestock genomics community. Investment in bioinformatics training capacity, HPC infrastructure, and international data sharing frameworks — particularly in Africa, South Asia, and Latin America where livestock genetic diversity is greatest and research capacity is most constrained — will be essential for ensuring equitable benefit from the NGS revolution in livestock science.

References

Daetwyler, H.D., Capitan, A., Pausch, H., et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46: 858–865.

Dong, Y., Xie, M., Jiang, Y., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat. Biotechnol. 31: 135–141.

Fontanesi, L. (2016). Nutrigenomics and nutrigenetics in livestock: Inputs from genomic tools and potential role in animal production. Anim. Front. 6: 18–26.

Groenen, M.A.M., Archibald, A.L., Uenishi, H., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491: 393–398.

Hess, M., Sczyrba, A., Egan, R., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463–467.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Koufariotis, L., Chen, Y.-P.P., Bolormaa, S., & Hayes, B.J. (2014). Regulatory and coding genome regions are enriched for trait-associated variants in dairy and beef cattle. BMC Genomics 15: 436.

Lemay, D.G., Lynn, D.J., Martin, W.F., et al. (2009). The bovine lactation genome: Insights into the evolution of mammalian milk. Genome Biol. 10: R43.

1000 Bull Genomes Project Consortium (2019). Whole-genome sequence association analysis boosts signal discovery in GWAS. PLoS Genet. 15: e1008500.

Rosen, B.D., Bickhart, D.M., Schnabel, R.D., et al. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9: giaa021.

Villar, D., Berthelot, C., Aldridge, S., et al. (2015). Enhancer evolution across 20 mammalian species. Cell 160: 554–566.

Warr, A., Affara, N., Aken, B., et al. (2020). An improved pig reference genome sequence to enable pig genetics and genomics research. Gigascience 9: giaa051.

Xue, Z., Huang, K., Cai, C., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500: 593–597.