

Article

Initial Analysis of Structural Variation Detections in Cattle Using Long-Read Sequencing Methods

Yahui Gao^{1,2}, Li Ma²  and George E. Liu^{1,*} 

¹ Animal Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705, USA; gyhalvin@gmail.com

² Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA; lima@umd.edu

* Correspondence: george.liu@usda.gov; Tel.: +1-301-504-9843

Abstract: Structural variations (SVs), as a great source of genetic variation, are widely distributed in the genome. SVs involve longer genomic sequences and potentially have stronger effects than SNPs, but they are not well captured by short-read sequencing owing to their size and relevance to repeats. Improved characterization of SVs can provide more advanced insight into complex traits. With the availability of long-read sequencing, it has become feasible to uncover the full range of SVs. Here, we sequenced one cattle individual using 10× Genomics (10 × G) linked read, Pacific Biosciences (PacBio) continuous long reads (CLR) and circular consensus sequencing (CCS), as well as Oxford Nanopore Technologies (ONT) PromethION. We evaluated the ability of various methods for SV detection. We identified 21,164 SVs, which amount to 186 Mb covering 7.07% of the whole genome. The number of SVs inferred from long-read-based inferences was greater than that from short reads. The PacBio CLR identified the most of large SVs and covered the most genomes. SVs called with PacBio CCS and ONT data showed high uniformity. The one with the most overlap with the results obtained by short-read data was PB CCS. Together, we found that long reads outperformed short reads in terms of SV detections.



Citation: Gao, Y.; Ma, L.; Liu, G.E. Initial Analysis of Structural Variation Detections in Cattle Using Long-Read Sequencing Methods. *Genes* **2022**, *13*, 828. <https://doi.org/10.3390/genes13050828>

Academic Editor: Qiuyue Liu

Received: 15 April 2022

Accepted: 4 May 2022

Published: 6 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: cattle; structural variation; long-read sequencing

1. Introduction

Unraveling the genetic underpinnings of phenotypic variation relies on a comprehensive knowledge of all forms of genetic variation. The exploitation of genetic variation has mainly focused on single-nucleotide polymorphisms (SNPs) and small insertions or deletions (indels, <50 bp), with a minor emphasis on larger variations such as copy number variations (CNV) and other structural variations (SV). SVs are most commonly defined as genomic changes of at least 50 bp in size, and they are difficult to detect precisely. Although there exist fewer SVs in the genome relative to SNPs and indels, SVs can impact more base pairs, thus being more likely to affect the phenotype [1,2]. While short-read sequencing technologies can detect SVs, they have various weaknesses. Since short reads (<1 kb) are typically smaller than or similar in size to SVs, a wide collection of indirect methods has been developed to infer SVs, including the use of split reads, read pairs, read depths, and local de novo assembly. On the other hand, linked reads provide long-range (100+ kb) information to short reads, bringing the reads into phase for haplotype-specific deletion detection, large SV detection [3–5], and diploid de novo assembly [6]. Long reads (>> 1 kb) spanning more SVs allow further SV detection, with mapped reads [7,8], local assembly after phasing long reads [9], and global de novo assembly [10,11]. Currently, Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are the most commonly employed technologies to produce long reads. Single-molecule real-time (SMRT) sequencing, developed by PacBio, can yield reads of tens of kilobases using either continuous long reads (CLR) or circular consensus sequencing (CCS) mode, which achieves high-quality

genome assembly. ONT enables direct and real-time sequencing of long DNA or RNA by analyzing the current interference caused by the molecules as they pass through the protein nanopore. To date, these sequencing methods have enabled the improved genome assemblies for many species, including humans [12,13], cattle [14–16], buffalo [17], pigs [18,19], sheep [20], and goats [21]. To study the effects of these methods on SV detection in humans, Aganezov et al. [22] performed whole-genome sequencing of the SKBR3 breast cancer cell line and patient-derived tumor and normal organoids from two breast cancer patients using Illumina/10× Genomics, PacBio, and ONT sequencing. They inferred SVs and large-scale CNVs and showed that long-read sequencing enables more accurate and sensitive SV detection. In dairy cattle, Couldrey et al. [15] detected CNVs using PacBio long-read and Illumina sequencing. In this study (Figure 1), we sequenced one cattle individual using cutting-edge technologies, i.e., 10× Genomics (10 × G), PromethION (ONT), PacBio continuous long reads (PB CLR), and PacBio circular consensus sequencing (PB CCS). We then evaluated various methods using these data from the same lung DNA sample for their abilities for the SV detection.

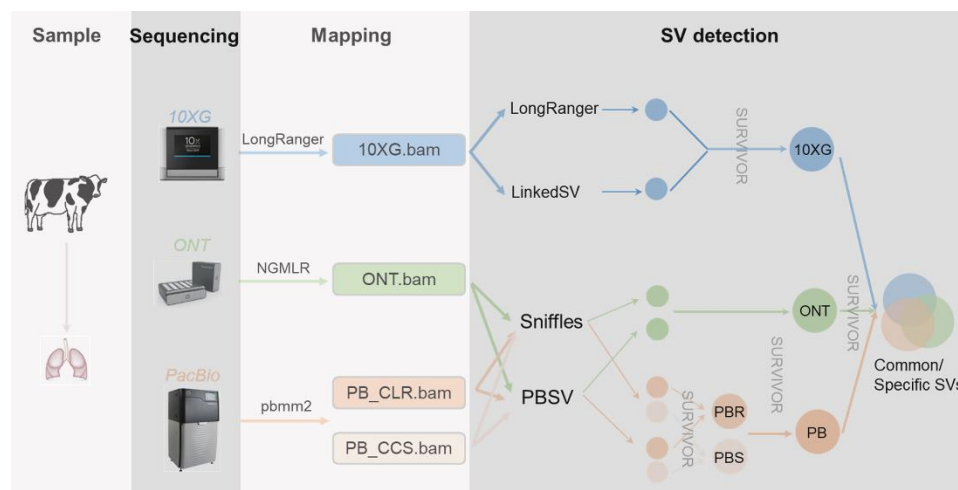


Figure 1. Sample collecting, sequencing, and mapping pipeline.

2. Materials and Methods

Under the approval of the US Department of Agriculture, Agricultural Research Service, Beltsville Agricultural Research Center’s Institutional Animal Care and Use Committee (Protocol 16-016), lung tissue was collected and then snap-frozen in liquid N₂ immediately after excision and kept at -80°C until use. The high-molecular-weight (HMW) DNA for lung tissue was extracted according to the MagAttract HMW DNA Kit (Cat. No. 67563, QIAGEN, Valencia, CA, USA). The quality of DNA samples was evaluated using the 2100 Bioanalyzer and the 4200 TapeStation (both from Agilent Technologies, Santa Clara, CA, USA), including degradation, potential RNA contamination, purity (OD₂₆₀/OD₂₈₀), and concentration using spectrophotometers of Qubit (Thermo Fisher Scientific, Waltham, MA, USA) and NanoDrop (NanoDrop Technologies, Rockland, DE, USA) to meet the demands for library construction.

The HMW DNA was sequenced using the Linked-Reads method developed by 10× Genomics [4], and standard protocols were followed in this study. By using microfluidics to segment and barcode HMW DNA, 10× Genomics can provide long-range information for short reads of the genome. We then aligned 10 × G short reads with LongRanger [23] v2.1.6 and used LongRanger [23] v 2.1.6 and LinkedSV [24] with the recommended settings to call SVs, respectively. DNA was prepared using standard ONT methods and sequenced on a PromethION device. We aligned ONT long reads with NGMLR [8] v0.2.7 and run Sniffles [8] v1.0.11 and PBSV v2.2.0 (<https://github.com/PacificBiosciences/pbsv>, accessed on 3 May 2022) with default settings for SV inference. PacBio sequencing was carried out on a Pacific Biosciences Sequel II platform using two modes, i.e., continuous long

reads (CLR) and circular consensus sequencing (CCS). We aligned the long reads with pbmm2 v1.3.0 (<https://github.com/PacificBiosciences/pbmm2>, accessed on 3 May 2022) and run Sniffles v1.0.11 [8] and PBSV v2.2.0 (<https://github.com/PacificBiosciences/pbsv>, accessed on 3 May 2022) with default settings for SV inference. We mapped all reads against the latest cattle genome reference ARS-UCD1.2 [25] and performed follow-up SV detection. We computed the alignment coverage by SAMtools [26] v1.9 depth command. For each sequencing technology, we merged the SVs generated by different callers with the SURVIVOR [27] v1.0.7 into a 10 × G, ONT, and PacBio technology-specific SV call sets. We then ran the SURVIVOR merge module with a maximum allowed distance of 200 bp and minimum SV size set to 30 bp regardless of SV types, as different methods may assign different types.

3. Results

3.1. SV Inference

A total of 1,577,259,728 (Table 1) short reads were generated through 10× Genomics, representing 55× coverage of the genome. The LongRanger alignment resulted in 97.14% (Table 1) of the reads mapping to the ARS-UCD1.2 cattle genome reference [25]. There was a total of 8315 and 6453 putative SVs identified by LongRanger and LinkedSV, respectively (Table 2). The SVs identified by LongRanger ranged in size from 49 bp to 1.59 Mb with an average size of 4481 bp (Table S1). For LinkedSV, the size ranged from 39 bp to 2.39 Mb, and the average size was 3180 bp (Table S1). The distribution of SVs across the genome was shown in Figure 2. After merging using SURVIVOR, the total quantity of SVs was 10,439 (114 duplications and 10,325 deletions) (Table 2), covering 53 Mb of the whole genome (Table S1).

Table 1. Yield and alignment coverage statistics for the cattle lung sample across various sequencing platforms.

Platform	10 × G	PromethION	PacBio CLR	PacBio CCS
Number of reads	1,577,259,728	1,618,623	11,178,388	2,875,796
Mapped reads	1,532,221,733	1,488,641	11,178,388	2,875,796
Mapping rate (%)	97.14	91.97	100	100
Depth	55×	11×	40×	6×
Read min length	19	70	53	74
Read max length	150	248,333	369,285	47,915
Read mean length	133.94	28,191.59	25,259.03	8763.78

Table 2. Statistics over SVs identified by various methods.

Platform	Method	DEL	DUP	Total
10 × G	LongRanger	8242	73	8315
	LinkedSV	6415	38	6453
	Merge	10,325	114	10,439
ONT	PBSV	26,397	2888	29,285
	Sniffles	3497	168	3665
	Merge	13,472	1881	15,353
PB_CLR	PBSV	885	169	1054
	Sniffles	1340	1238	2578
	Merge	1800	1162	2962
PB_CCS	PBSV	23,353	6569	29,922
	Sniffles	190	99	289
	Merge	15,601	3891	19,492
Merge	SURVIVOR	16,289	4875	21,164



Figure 2. Individualized cattle SV map. The tracks under every black bar represent the SVs for $10 \times$ G_LongRanger, $10 \times$ G_LinkedSV, CCS_PBSV, CCS_Sniffles, CLR_PBSV, CLR_Sniffles, ONT_PBSV and ONT_Sniffles (in order from top to bottom). Red means deletion, and green means duplication.

Oxford nanopore sequencing generated 1,618,623 sequences representing approximately $11\times$ coverage of the genome (Table 1). The distribution of sequence lengths (70–248,333 bp) was shown in Figure S1a, with an average length of 28,191.59 bp (Table 1). A total of 91.97% (Table 1) of the reads were mapped to the cattle genome assembly. Sniffles and PBSV identified 3665 and 29,285 SVs, respectively (Table 2). The identified SVs ranged from 32 bp to 2.62 Mb (mean size = 5676 bp) and 9 bp to 0.1 Mb (mean size = 592 bp) (Table S1), and their distribution across the whole genome was shown in Figure 2. The merging total number of SVs was 15,353 (1881 duplications and 13,472 deletions) (Table 2), covering 34 Mb of the whole genome (Table S1).

PacBio CLR sequencing yielded a total of 11,178,388 reads, representing 40-fold genome coverage, and they distributed in length between 53 and 369,285 bp (Figure S1b), with an average of 25,259.03 bp (Table 1). All reads were mapped to the cattle reference genome by pbmm2 (Table 1). Sniffles and PBSV identified 2578 and 1054 SVs, respectively (Table 2). The SV sizes identified by Sniffles ranged from 35 bp to 2.62 Mb, with an average size of 36,485 bp (Figure 2 and Table S1). For PBSV, the sizes ranged from 14 bp to 96 kb, and the mean size was 2377 bp (Figure 2 and Table S1). A total 2962 (1162 duplications and 1800 deletions) events covering 92 Mb (Table S1) of the whole genome were identified after merging (Table 2).

PacBio CCS sequencing generated 2,875,796 reads, representing $6\times$ coverage of the genome. The distribution of sequence length (74–47,915 bp) is illustrated in Figure S1c, with an average length of 8763.78 bp (Table 1). All reads were mapped to the cattle reference genome by pbmm2 (Table 1). Sniffles and PBSV identified 289 and 29,922 putative SVs, respectively (Table 2). The SV sizes identified by Sniffles ranged from 34 bp to 3.6 Mb and had a mean size of 72,166 bp (Figure 2, Table S1). For PBSV, the sizes ranged from 8 bp to 100 kb, and the mean size was 722 bp (Figure 2 and Table S1). The total merging number of SVs was 19,492 (3891 duplications and 15,601 deletions) (Table 2), covering 41 Mb of the whole genome (Table S1).

3.2. SV Overlap

In general, the total amount of SVs derived from short reads is much smaller than for the long-read-based inferences (Table 2). Most of the SVs were located between 50 bp to 200 bp, but long-read-based inferences can detect more large SVs (Figure 3a). Overall, these results show that across SVs accounts and sizes, long-read-based SV inference outperforms that of short reads. Between 45% and 60% of variants were called in at least one of the long-read data types, both of which were supported (Figure 3b). SVs called using PacBio CCS and ONT data showed high concordance (Figure 3b). The highest overlap with the results obtained from the short-read data was the PacBio CCS.

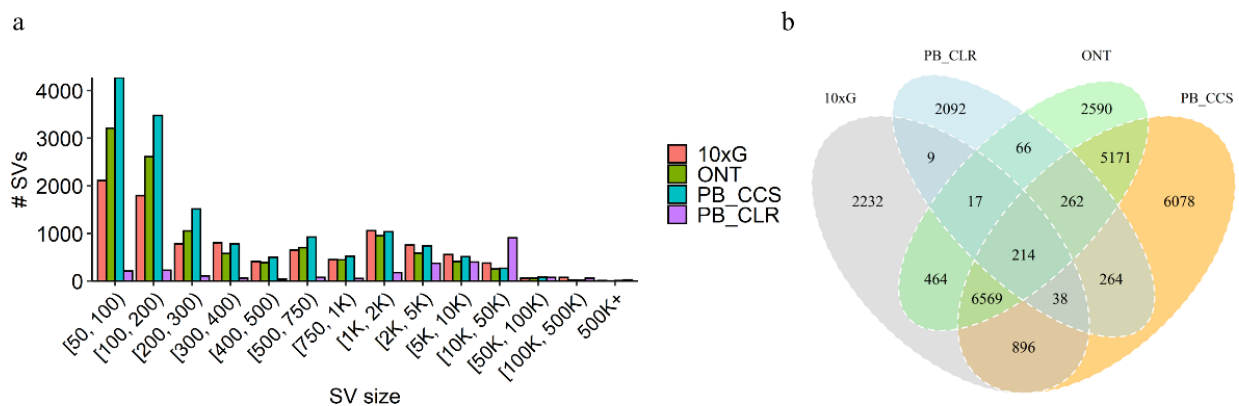


Figure 3. (a) Size distribution for SVs inferred from either long reads or Illumina/10 × G short reads. (b) Comparison between the four SV datasets.

4. Discussion

The long reads generated by the third-generation sequencing technology can span tens of thousands of base pairs, which are tremendously serviceable in filling gaps in current references [28,29] and for the assembly of complicated genomic regions [29,30]. Meanwhile, they can also be helpful for the identification of large SVs. In this study, we presented a comparison of four sequencing datasets from the same cattle lung DNA sample. We sequenced the genome with Illumina/10 × G, ONT, and PacBio (CLR and CCS) sequencing technologies and subsequently analyzed for structural variations. We observed comparisons between various SV methods and how SV results differ for different sequencing technologies.

We identified a total of 21,164 SVs, which amount to 186 Mb covering 7.07% of the whole genome (Table 2). In general, except for PB CLRs, the number of SVs inferred from long-read-based inferences was greater than that of short-reads (Table 2). The CLR detected the least number of SVs, probably due to insufficient coverage, but it identified the most of large SVs and covered the most genomes (Figure 3a). When using 10× linked reads, we obtained 10,439 SVs, but there were 8207 SVs shared between short- and long-read technologies (Figure 3a). We showed that SVs called with PacBio CCS versus ONT data show high concordance, with more than 90% of SVs called with one platform also being called with the other (Figure 3a), which is consistent with human results [22]. Our results indicated a concordance between SVs inferred with ONT and PacBio CCS.

With the advancement of long-read sequencing, the higher quality of the reference assembly could further benefit the identification of SVs. Leonard et al. showed that 20× for HiFi or 60× for ONT sequencing was sufficient to produce two haplotype-resolved assemblies while retaining over 90% accuracy in detecting SVs when integrated into pangenomes [31]. With a combination of PacBio HiFi, Hi-C, and ONT ultra-long read sequencing, we could soon routinely obtain a Telomere-to-Telomere (T2T) assembly for livestock, as recently demonstrated for humans [32].

5. Conclusions

In this study, we generated four sequencing datasets and compared the SV results based on them. For each dataset, we identified SVs using two programs. Our results indicated a concordance between SVs inferred with ONT and PacBio CCS. The one with the most overlap with the results obtained by short-read data is PB CCS. Together, we found that long reads performed better than short reads in terms of SV detections.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13050828/s1>. Figure S1: Length distribution for reads from ONT and PacBio sequencing runs. Table S1. Summary of identified SVs using different methods.

Author Contributions: Conceptualization, Y.G. and G.E.L.; methodology, Y.G.; software, Y.G.; formal analysis, Y.G.; resources, L.M.; writing—original draft preparation, Y.G.; writing—review and editing, G.E.L.; supervision, G.E.L.; funding acquisition, L.M. and G.E.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by USDA National Institute of Food and Agriculture (NIFA) Agriculture and Food Research Initiative (AFRI) grant numbers 2016-67015-24886, 2019-67015-29321, and 2021-67015-33409.

Institutional Review Board Statement: The animal study protocol was approved by the US Department of Agriculture, Agricultural Research Service, Beltsville Agricultural Research Center's Institutional Animal Care and Use Committee (Protocol 16-016).

Informed Consent Statement: Not applicable.

Data Availability Statement: The SV calls reported in this article are available in the Supplemental Material, and sequencing data are available upon request for research purposes.

Acknowledgments: We thank Reuben Anderson, Mary Bowman, Donald Carbaugh, Christina Clover, Sarah McQueeney, Mary Niland, Derek Bickhart, Tim Smith, and Research Animal Services for the technical assistance. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture (USDA). The USDA is an equal opportunity provider and employer.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Chaisson, M.; Sanders, A.D.; Zhao, X.; Malhotra, A.; Porubsky, D.; Rausch, T.; Gardner, E.J.; Rodriguez, O.L.; Guo, L.; Collins, R.L.; et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* **2019**, *10*, 1784. [[CrossRef](#)] [[PubMed](#)]
2. Chiang, C.; Scott, A.J.; Davis, J.R.; Tsang, E.K.; Li, X.; Kim, Y.; Hadzic, T.; Damani, F.N.; Ganel, L.; GTEx Consortium; et al. The impact of structural variation on human gene expression. *Nat. Genet.* **2017**, *49*, 692–699. [[CrossRef](#)] [[PubMed](#)]
3. Karaoglanoglu, F.; Ricketts, C.; Ebre, E.; Rasekh, M.E.; Hajirasouliha, I.; Alkan, C. VALOR2: Characterization of large-scale structural variants using linked-reads. *Genome Biol.* **2020**, *21*, 72. [[CrossRef](#)] [[PubMed](#)]
4. Marks, P.; Garcia, S.; Barrio, A.M.; Belhocine, K.; Bernate, J.; Bharadwaj, R.; Bjornson, K.; Catalanotti, C.; Delaney, J.; Fehr, A.; et al. Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Res.* **2019**, *29*, 635–645. [[CrossRef](#)] [[PubMed](#)]
5. Spies, N.; Weng, Z.; Bishara, A.; McDaniel, J.; Catoe, D.; Zook, J.M.; Salit, M.; West, R.B.; Batzoglou, S.; Sidow, A. Genome-wide reconstruction of complex structural variants using read clouds. *Nat. Methods* **2017**, *14*, 915–920. [[CrossRef](#)] [[PubMed](#)]
6. Weisenfeld, N.I.; Kumar, V.; Shah, P.; Church, D.M.; Jaffe, D.B. Direct determination of diploid genome sequences. *Genome Res.* **2017**, *27*, 757–767. [[CrossRef](#)] [[PubMed](#)]
7. Cretu Stancu, M.; van Roosmalen, M.J.; Renkens, I.; Nieboer, M.M.; Middelkamp, S.; de Ligt, J.; Pregno, G.; Giachino, D.; Mandrile, G.; Espejo Valle-Inclan, J.; et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **2017**, *8*, 1326. [[CrossRef](#)]
8. Sedlazeck, F.J.; Rescheneder, P.; Smolka, M.; Fang, H.; Nattestad, M.; von Haeseler, A.; Schatz, M.C. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **2018**, *15*, 461–468. [[CrossRef](#)]
9. Chaisson, M.J.; Huddleston, J.; Dennis, M.Y.; Sudmant, P.H.; Malig, M.; Hormozdiari, F.; Antonacci, F.; Surti, U.; Sandstrom, R.; Boitano, M.; et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **2015**, *517*, 608–611. [[CrossRef](#)]
10. 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.; et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [[CrossRef](#)]
11. Koren, S.; Rhie, A.; Walenz, B.P.; Dilthey, A.T.; Bickhart, D.M.; Kingan, S.B.; Hiendleder, S.; Williams, J.L.; Smith, T.; Phillippy, A.M. De novo assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* **2018**, *36*, 174–1182. [[CrossRef](#)] [[PubMed](#)]
12. Ebert, P.; Audano, P.A.; Zhu, Q.; Rodriguez-Martin, B.; Porubsky, D.; Bonder, M.J.; Sulovari, A.; Ebler, J.; Zhou, W.; Serra Mari, R.; et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **2021**, *372*, eabf7117. [[CrossRef](#)] [[PubMed](#)]
13. Quan, C.; Li, Y.; Liu, X.; Wang, Y.; Ping, J.; Lu, Y.; Zhou, G. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol.* **2021**, *22*, 159. [[CrossRef](#)]
14. Low, W.Y.; Tearle, R.; Liu, R.; Koren, S.; Rhie, A.; Bickhart, D.M.; Rosen, B.D.; Kronenberg, Z.N.; Kingan, S.B.; Tseng, E.; et al. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat. Commun.* **2020**, *11*, 2071. [[CrossRef](#)]
15. Couldrey, C.; Keehan, M.; Johnson, T.; Tiplady, K.; Winkelman, A.; Littlejohn, M.D.; Scott, A.; Kemper, K.E.; Hayes, B.; Davis, S.R.; et al. Detection and assessment of copy number variation using PacBio long-read and Illumina sequencing in New Zealand dairy cattle. *J. Dairy Sci.* **2017**, *100*, 5472–5478. [[CrossRef](#)] [[PubMed](#)]
16. Lamb, H.J.; Ross, E.M.; Nguyen, L.T.; Lyons, R.E.; Moore, S.S.; Hayes, B.J. Characterization of the poll allele in Brahman cattle using long-read Oxford Nanopore sequencing. *J. Anim. Sci.* **2020**, *98*, skaa127. [[CrossRef](#)]
17. Ananthasayanam, S.; Kothandaraman, H.; Nayee, N.; Saha, S.; Baghel, D.S.; Gopalakrishnan, K.; Peddamma, S.; Singh, R.B.; Schatz, M. First near complete haplotype phased genome assembly of River buffalo (*Bubalus bubalis*). *bioRxiv* **2020**, 618785. [[CrossRef](#)]
18. Zhou, R.; Li, S.T.; Yao, W.Y.; Xie, C.D.; Chen, Z.; Zeng, Z.J.; Wang, D.; Xu, K.; Shen, Z.J.; Mu, Y.; et al. The Meishan pig genome reveals structural variation-mediated gene expression and phenotypic divergence underlying Asian pig domestication. *Mol. Ecol. Resour.* **2021**, *21*, 2077–2092. [[CrossRef](#)] [[PubMed](#)]
19. Ma, H.; Jiang, J.; He, J.; Liu, H.; Han, L.; Gong, Y.; Li, B.; Yu, Z.; Tang, S.; Zhang, Y.; et al. Long-read assembly of the Chinese indigenous Ningxiang pig genome and identification of genetic variations in fat metabolism among different breeds. *Mol. Ecol. Resour.* **2022**, *22*, 1508–1520. [[CrossRef](#)]

20. Li, R.; Gong, M.; Zhang, X.; Wang, F.; Liu, Z.; Zhang, L.; Xu, M.; Zhang, Y.; Dai, X.; Zhang, Z.; et al. The first sheep graph-based pan-genome 1 reveals the spectrum of structural variations and their effects on tail phenotypes. *bioRxiv* **2021**, 472709. [[CrossRef](#)]
21. Bickhart, D.M.; Rosen, B.D.; Koren, S.; Sayre, B.L.; Hastie, A.R.; Chan, S.; Lee, J.; Lam, E.T.; Liachko, I.; Sullivan, S.T.; et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **2017**, *49*, 643–650. [[CrossRef](#)] [[PubMed](#)]
22. Aganezov, S.; Goodwin, S.; Sherman, R.M.; Sedlazeck, F.J.; Arun, G.; Bhatia, S.; Lee, I.; Kirsche, M.; Wappel, R.; Kramer, M.; et al. Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing. *Genome Res.* **2020**, *30*, 1258–1273. [[CrossRef](#)] [[PubMed](#)]
23. Zheng, G.X.; Lau, B.T.; Schnall-Levin, M.; Jarosz, M.; Bell, J.M.; Hindson, C.M.; Kyriazopoulou-Panagiotopoulou, S.; Masquelier, D.A.; Merrill, L.; Terry, J.M.; et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **2016**, *34*, 303–311. [[CrossRef](#)] [[PubMed](#)]
24. Fang, L.; Kao, C.; Gonzalez, M.V.; Mafra, F.A.; Pellegrino da Silva, R.; Li, M.; Wenzel, S.S.; Wimmer, K.; Hakonarson, H.; Wang, K. LinkedSV for detection of mosaic structural variants from linked-read exome and genome sequencing data. *Nat. Commun.* **2019**, *10*, 5585. [[CrossRef](#)]
25. Rosen, B.D.; Bickhart, D.M.; Schnabel, R.D.; Koren, S.; Elisk, C.G.; Tseng, E.; Rowan, T.N.; Low, W.Y.; Zimin, A.; Couldrey, C.; et al. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **2020**, *9*, giaa021. [[CrossRef](#)]
26. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
27. Jeffares, D.C.; Jolly, C.; Hoti, M.; Speed, D.; Shaw, L.; Rallis, C.; Balloux, F.; Dessimoz, C.; Bähler, J.; Sedlazeck, F.J. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **2017**, *8*, 14061. [[CrossRef](#)]
28. English, A.C.; Richards, S.; Han, Y.; Wang, M.; Vee, V.; Qu, J.; Qin, X.; Muzny, D.M.; Reid, J.G.; Worley, K.C.; et al. Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **2012**, *7*, e47768. [[CrossRef](#)]
29. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)]
30. English, A.C.; Salerno, W.J.; Hampton, O.A.; Gonzaga-Jauregui, C.; Ambreth, S.; Ritter, D.I.; Beck, C.R.; Davis, C.F.; Dahdouli, M.; Ma, S.; et al. Assessing structural variation in a personal genome-towards a human reference diploid genome. *BMC Genom.* **2015**, *16*, 286. [[CrossRef](#)]
31. Leonard, A.S.; Crysanto, D.; Fang, Z.H.; Heaton, M.P.; Ley, B.L.V.; Herrera, C.; Bollwein, H.; Bickhart, D.M.; Kuhn, K.L.; Smith, T.P.L.; et al. Structural variant-based pangenome construction has low sensitivity to variability of haplotype-resolved bovine assemblies. *bioRxiv* **2021**, 466900. [[CrossRef](#)]
32. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The complete sequence of a human genome. *Science* **2022**, *376*, 44–53. [[CrossRef](#)] [[PubMed](#)]