Short communication
DOI: /10.5513/JCEA01/20.4.2679
Journal of Central European Agriculture, 2019, 20(4), p.1111-1115

The new bovine reference genome assembly provides new insight into genomic organization of the bovine major histocompatibility complex

Minja ZORC, Jernej OGOREVC, Peter DOVČ (✉)

University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Jamnikarjeva 101 SI-1000 Ljubljana, Slovenia

✉ Corresponding author: peter.dovc@bf.uni-lj.si

ABSTRACT

The reference genome sequence represents a key resource for genetic studies of the target species. In 2009, two reference assemblies of the cattle (*Bos taurus*) genome, were published (Btau 4.0 and UMD 2.0). Both assemblies were upgraded several times since then. Highly polymorphic major histocompatibility complex (MHC) encodes proteins crucial for immune recognition and regulation of immune response in vertebrates. It is characterised by extensive nucleotide diversity, copy number variation of paralogous genes, and long repetitive sequences. In cattle, MHC is designated as BoLA (bovine leucocyte antigen), located on the chromosome 23. Its organisation differs from typical mammalian MHCs. The structural complexity makes it difficult to assemble a reliable reference sequence of this genomic region. Therefore, this region represents a good genomic model region to compare the accuracy of different assembly strategies. Recent advances in long-read sequence technology, combined with new scaffolding technologies, enabled issuing of the new bovine reference genome assembly build ARS-UCD 1.2, which is significantly improved over previous bovine genome assembly releases. In the current study the software tool Mauve for multiple alignment of conserved genomic sequences with rearrangements was used to identify the differences of genomic organization in the BoLA region assembled in three bovine reference genomes, Btau 5.0.1, UMD 3.1.1, and ARS-UCD 1.2. Multiple alignment of the bovine chromosome 23 sequences extracted from three genome assemblies revealed differences in the structure of the BoLA region. Segments encoding genes *BOLA-DMA* and *BOLA-DQB* are rearranged and inverted in the new assembly relative to the previous builds.

Keywords: BoLA, *Bos taurus*, genome assembly, optical map, reference genome

INTRODUCTION

A reference genome sequence assembly represents a key resource for genetic studies of the target species. The field of genome assembly is constantly evolving (Steinberg et al., 2017). Due to the economic importance of domestic cattle (*Bos taurus*) and its fascinating abilities e.g. lactation and comprehensive immune system, it became the first mammalian livestock species to have its genome completely sequenced. The sequencing and assembly of the bovine genome began at Baylor College of Medicine in 2003 and has been completed in 2009 with publishing of Btau 4.0 assembly (Elsik et al., 2009), representing combination of individual genomes of the Limousine bull Domino (BAC path) and his daughter Dominette (9x coverage whole genome sequence (WGS)). An alternative assembly of the bovine genome was built at the University of Maryland (UMD 2.0) using the same set of sequencing data generated for Baylor assembly but different software and assembly approach (Zimin et al., 2009). There were significant inconsistencies between the two assemblies (Zimin et al., 2012), which were both

upgraded several times since their first release. The existence of two different reference genome assemblies impacts their use in genomic studies, including search for candidate genes for certain traits and genomic selection. The main differences distinguishing both bovine genome assemblies are more unassigned sequences to chromosomes present in Btau than in UMD (Partipilo et al., 2011) and the presence of Y chromosome in Btau assembly only.

A comparison of two reference assemblies (Btau 4.1 and UMD 3.1) with other sequencing data revealed significant differences especially in genomic regions with repetitive elements, such as number and size of some major histocompatibility complex (MHC) class I pseudogenes and intergenic regions (Schwartz and Hammond, 2015). The differences between MHC class II genome sequence and linkage maps were also reported (Takeshima and Aida, 2006). The MHC is one of the most dynamic genome regions. It is divided into three subregions (classes I, II, III) encoding genes of similar functions but different structures among species (Flajnik et al., 1999; Kulski et al., 2002). As species evolved and adapted to pathogenic pressures, the gene content and orientation of MHC regions has diverged (Kelley et al., 2005). It is characterized by extensive nucleotide diversity, copy number variation of paralogous genes, and long repetitive sequences. In cattle, MHC is known as the bovine leukocyte antigen (BoLA), and is located on autosome 23 (Spooner et al., 1978) and organized differently from typical mammalian MHCs (Childers et al., 2006). The large genetic distance, spanning at least 15 cM between the two subregions of the BoLA class II (classes IIa and IIb), distinguishes the bovine MHC from those of human and mouse (Andersson et al., 1988).

Recent advances in long-read sequence technology, combined with new scaffolding technologies, enabled the new bovine reference genome assembly build. The new bovine assembly, ARS-UCD 1.2, was released in April 2018. PacBio sequences of an approximately 80x genome coverage were assembled *de novo*. Scaffolding of the novel assembly is based on the Dovetail Genomics Chicago data, BtOM1.0 optical map (Zhou et al., 2015) and recombination map of 59K autosomal SNPs. Additionally, full-length transcripts from 28 cow tissues were sequenced with PacBio using the Iso-Seq method to support improved genome annotation. According to the new assembly statistics, there are significant improvements over UMD3.1 release. The aim of the present study was comparison of the MHC region between new bovine assembly ARS-UCD 1.2 with previous two assemblies, Btau 5.0.1 and UMD 3.1.1.

MATERIALS AND METHODS

Three bovine reference genomes (Btau 5.0.1, UMD 3.1.1, and ARS-UCD 1.2) were downloaded from GenBank server. The sequence of chromosome 23 in GenBank format which contains both, sequences and annotations, was used in the analysis. The tool Mauve (Darling et al., 2004) for multiple alignment of conserved genomic sequences with rearrangements was used to identify the differences of the genomic organization of the chromosome 23 that contains BoLA region. In the analysis progressiveMauve algorithm was applied (Darling et al., 2010)

RESULTS AND DISCUSSION

The multiple alignment of the bovine chromosome 23 sequences extracted from three genome assemblies revealed differences in the structure of the BoLA region (Figure 1). Each genome assembly is positioned horizontally. Homologous segments are presented as coloured blocks connected across assemblies. Downward shifted blocks represent segments inverted relative to the genome selected as reference (UMD 3.1.1).

Segments encoding genes *BOLA-DMA* (Figure 2) and *BOLA-DQB* (Figure 3) are rearranged and inverted in the new assembly, relative to previous builds.

The comparison of the three bovine genome assemblies revealed structural differences among them in the BoLA region at the bovine chromosome 23. The BoLA region is characterised by numerous duplications and pseudogenes including repetitive sequence elements,

which makes accurate assembly of this region difficult. The introduction of long genomic sequences allows, in spite of relatively high rate of sequencing errors, more efficient and accurate assembling. The new assembly, ARS-UCD 1.2 resolves some conflicts between the two old assemblies (Btau 5.0.1 and UMD 3.1.1) and some other BoLA targeted sequencing data, which improves the usefulness of the new bovine reference genome for functional genomics studies and genomic selection.

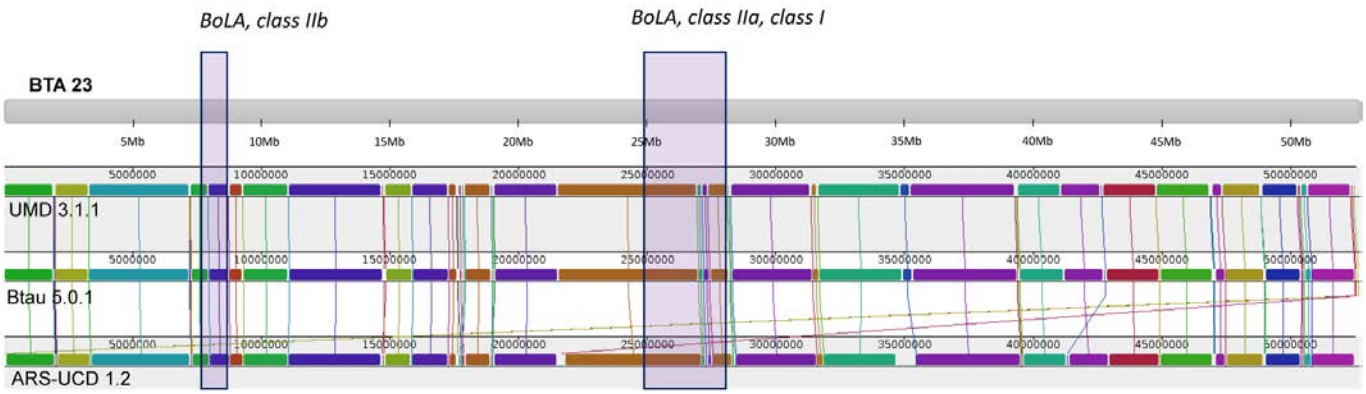

Figure 1. The multiple alignment of the bovine chromosome 23 sequences extracted from three genome assemblies revealed differences in the structure of the BoLA region assembled in three different genome assembly builds

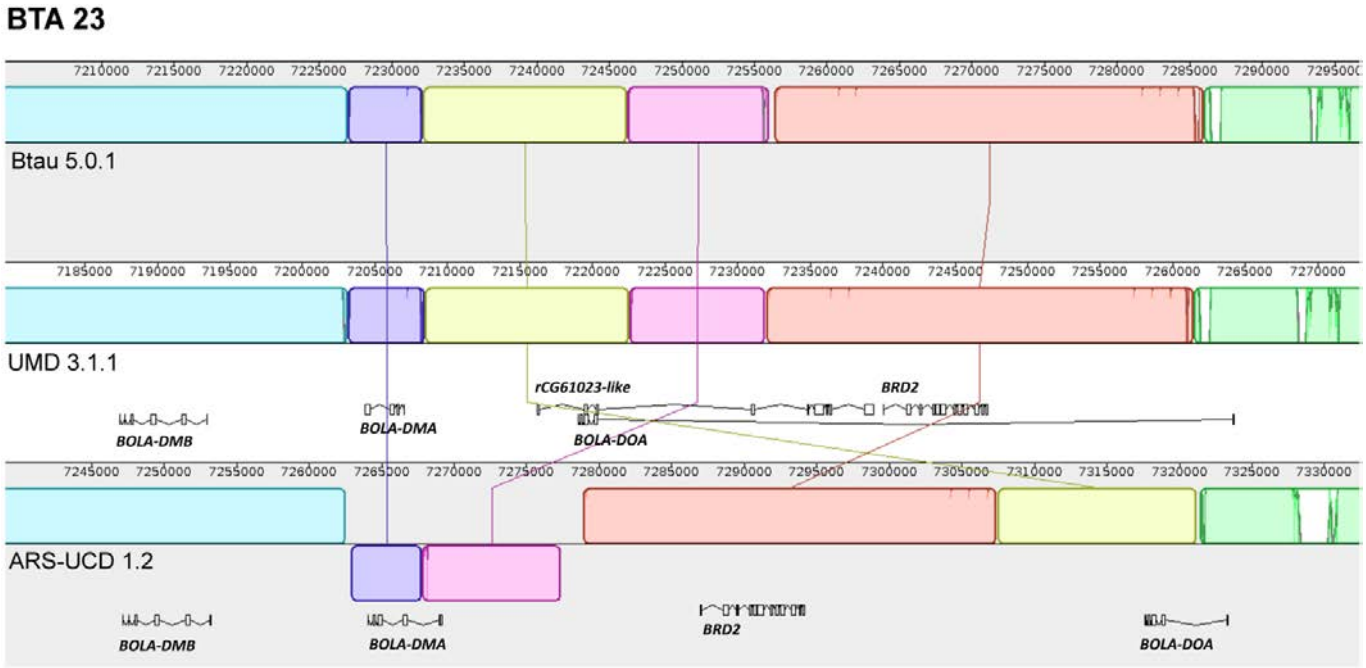

Figure 2. BoLA class IIb. Segment encoding *BOLA-DMA* is rearranged and inverted in the new assembly relative to the previous builds

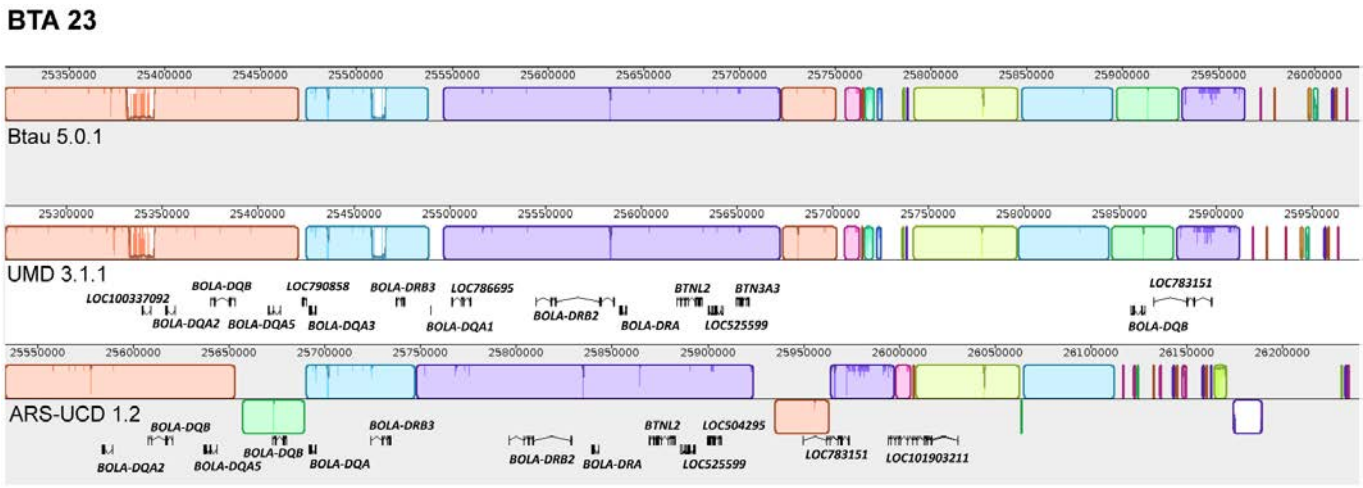

Figure 3. BoLA class IIa, class I. Segment encoding *BOLA-DQB* is rearranged and inverted in the new assembly relative to the previous builds

CONCLUSIONS

The introduction of novel technologies (e.g. optical mapping and long-range sequencing) represent significant advantage over the old BAC- and short WGS reads-based assembly strategies. The novel strategy is especially efficient for assembling of genomic regions with inversions, duplications and translocations.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding P4-0220: Comparative genomics and genomic biodiversity).

REFERENCES

Andersson, L., Lundén, A., Sigurdardottir, S., Davies, C.J., Rask, L. (1988) Linkage relationships in the bovine MHC region. High recombination frequency between class II subregions. Immunogenetics, 27 (4), 273-280. DOI: https://dx.doi.org/10.1007/BF00376122

Childers, C.P., Newkirk, H.L., Honeycutt, D.A., Ramlachan, N., Muzney, D.M., Sodergren, E., Gibbs, R.A., Weinstock, G.M., Womack, J.E., Skow, L.C. (2006) Comparative analysis of the bovine MHC class IIb sequence identifies inversion breakpoints and three unexpected genes. Animal Genetics, 37 (2), 121-129. DOI: https://dx.doi.org/10.1111/j.1365-2052.2005.01395.x

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Research, 14 (7), 1394-1403. DOI: https://dx.doi.org/10.1101/gr.2289704

Darling, A.E., Mau, B., Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One, 5 (6), e11147. DOI: https://dx.doi.org/10.1371/journal.pone.0011147

Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Consortium B.G.S.a.A. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science, 324 (5926), 522-528. DOI: https://dx.doi.org/10.1126/science.1169588

Flajnik, M.F., Ohta, Y., Namikawa-Yamada, C., Nonaka, M. (1999) Insight into the primordial MHC from studies in ectothermic vertebrates. Immunological Reviews, 167 (1), 59-67. DOI: https://dx.doi.org/10.1111/j.1600-065X.1999.tb01382.x

Kelley, J., Walter, L., Trowsdale, J. (2005) Comparative genomics of major histocompatibility complexes. Immunogenetics, 56 (10), 683-695. DOI: https://dx.doi.org/10.1007/s00251-004-0717-7

Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., Inoko, H. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunological Reviews, 190, 95-122. DOI: https://dx.doi.org/10.1034/j.1600-065X.2002.19008.x

Partipilo, G., D'Addabbo, P., Lacalandra, G.M., Liu, G.E., Rocchi, M. (2011) Refinement of Bos taurus sequence assembly based on BAC-FISH experiments. BMC Genomics, 12, 639-648. DOI: https://dx.doi.org/10.1186/1471-2164-12-639

Schwartz, J.C., Hammond, J.A. (2015) The assembly and characterisation of two structurally distinct cattle MHC class I haplotypes point to the mechanisms driving diversity. Immunogenetics, 67 (9), 539-544. DOI: https://dx.doi.org/10.1007/s00251-015-0859-9

Spooner, R.L., Leveziel, H., Grosclaude, F., Oliver, R.A., Vaiman, M. (1978) Evidence for a possible major histocompatibility complex (BLA) in cattle. International Journal of Immunogenetics, 5 (5), 325-546. DOI: https://dx.doi.org/10.1111/j.1744-313X.1978.tb00662.x

Steinberg, K.M., Schneider, V.A., Alkan, C., Montague, M.J., Warren, W.C., Church, D.M., Wilson, R.K. (2017) Building and improving reference genome assemblies. Proceedings of the IEEE, 105 (3), 422-435. DOI: https://dx.doi.org/10.1109/JPROC.2016.2645402

Takeshima, S.N., Aida, Y. (2006) Structure, function and disease susceptibility of the bovine major histocompatibility complex. Animal Science Journal, 77 (2), 138-150. DOI: https://dx.doi.org/10.1111/j.1740-0929.2006.00332.x

Zhou, S., Goldstein, S., Place, M., Bechner, M., Patino, D., Potamousis, K., Ravindran, P., Pape, L., Rincon, G., Hernandez-Ortiz, J., Medrano, J.F., Schwartz, D.C. (2015) A clone-free, single molecule map of the domestic cow (Bos taurus) genome. BMC Genomics, 16 (1), 644. DOI: https://dx.doi.org/10.1186/s12864-015-1823-7

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marcais, G., Roberts, M., Subramanian, P., Yorke, J.A., Salzberg, S.L. (2009) A whole-genome assembly of the domestic cow, Bos taurus. Genome Biology, 10 (4), R42. DOI: https://dx.doi.org/10.1186/gb-2009-10-4-r42

Zimin, A.V., Kelley, D.R., Roberts, M., Marçais, G., Salzberg, S.L., Yorke, J.A. (2012) Mis-assembled "segmental duplications" in two versions of the Bos taurus genome. PLoS ONE, 7 (8), e42680. DOI: https://dx.doi.org/10.1371/journal.pone.0042680