

# Bovine Genome Database: new annotation tools for a new reference genome

Md Shamimuzzaman<sup>1</sup>, Justin J. Le Tourneau<sup>1</sup>, Deepak R. Unni<sup>2</sup>, Colin M. Diesh<sup>3</sup>, Deborah A. Triant<sup>1</sup>, Amy T. Walsh<sup>1</sup>, Aditi Tayal<sup>1</sup>, Gavin C. Conant<sup>4,5,6</sup>, Darren E. Hagen<sup>7</sup> and Christine G. Elsik<sup>1,8,9,\*</sup>

<sup>1</sup>Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA, <sup>2</sup>Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94608, USA, <sup>3</sup>Department of Bioengineering, University of California, Berkeley, CA 94720, USA, <sup>4</sup>Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, USA, <sup>5</sup>Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA, <sup>6</sup>Program in Genetics, North Carolina State University, Raleigh, NC 27695, USA, <sup>7</sup>Department of Animal and Food Sciences, Oklahoma State University, Stillwater, OK 74078, USA, <sup>8</sup>Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA and <sup>9</sup>MU Institute for Data Science and Informatics, University of Missouri, Columbia, MO 65211, USA

Received September 13, 2019; Revised October 07, 2019; Editorial Decision October 08, 2019; Accepted October 08, 2019

## ABSTRACT

The Bovine Genome Database (BGD) (<http://bovinegenome.org>) has been the key community bovine genomics database for more than a decade. To accommodate the increasing amount and complexity of bovine genomics data, BGD continues to advance its practices in data acquisition, curation, integration and efficient data retrieval. BGD provides tools for genome browsing (JBrowse), genome annotation (Apollo), data mining (BovineMine) and sequence database searching (BLAST). To augment the BGD genome annotation capabilities, we have developed a new Apollo plug-in, called the Locus-Specific Alternate Assembly (LSAA) tool, which enables users to identify and report potential genome assembly errors and structural variants. BGD now hosts both the newest bovine reference genome assembly, ARS-UCD1.2, as well as the previous reference genome, UMD3.1.1, with cross-genome navigation and queries supported in JBrowse and BovineMine, respectively. Other notable enhancements to BovineMine include the incorporation of genomes and gene annotation datasets for non-bovine ruminant species (goat and sheep), support for multiple assemblies per organism in the Regions Search tool, integration of additional ontologies and development of many new template queries. To better serve the research community, we continue to focus on improving existing tools, developing new tools, adding new

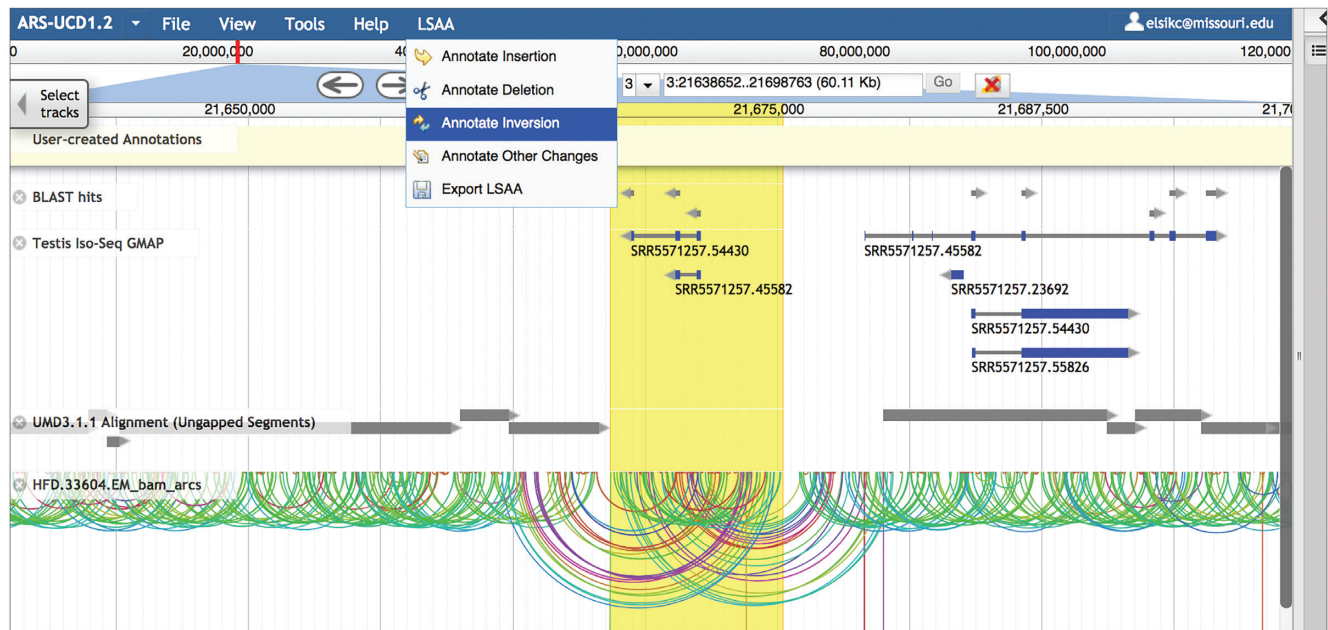
datasets and encouraging researchers to use these resources.

## INTRODUCTION

The availability of multiple livestock reference genome sequences has accelerated genomics based livestock research and breeding programs (1). The Bovine Genome Database (BGD) (2) provides easy access to the bovine genome assemblies and annotations via graphical genome browsers (JBrowse) (3), annotation tools (Apollo) (4), sequence search (BLAST) (5,6) and data mining (BovineMine) (2,7). Genome browsers at BGD provide different genomic features such as protein-coding and non-coding genes from RefSeq (8) and Ensembl (9), single nucleotide polymorphisms (SNP) from Ensembl Variation (9), quantitative trait loci (QTL) from AnimalQTLdb (10), as well as expression tracks computed using publicly available RNA-seq data (NCBI Sequence Read Archive (11)). BovineMine integrates these data with other external sources, including protein annotation (UniProt (12)), protein families and domains (InterPro (13)), orthologs and paralogs (EnsemblCompara (14), OrthoDB (15), TreeFam (16)), pathways (KEGG (17), Reactome (18)), interactions (BioGRID (19), IntAct (20)), Gene Ontology annotations (UniProt GOA (21)), SNP alias ids (SNPChIMp (22)) and publications (PubMed (11)).

BGD was first published in *Nucleic Acids Research* in 2011 (23), with the initial goal of supporting the Bovine Genome Sequencing and Analysis Consortium (BGSAC) (24) by providing web-based tools for annotation of the bovine genome. An update in the NAR database issue five years later (2) reported new genome browsing and anno-

\*To whom correspondence should be addressed. Tel: +1 573 884 7422; Fax: +1 573 882 6527; Email: elsikc@missouri.edu



**Figure 1.** Apollo browser showing the LSAA pull-down menu and evidence for a genome assembly inversion. The *BLAST hits* track shows that segments of a single cDNA align to the genome in opposite directions and the *Testis Iso-Seq GMAP* track shows that some full-length Iso-Seq transcripts have partial alignments in opposite directions. The *UMD3.1.1 Alignment (Ungapped Segments)* track reveals a region without an alternate assembly alignment. In the *HFD.33604.EM.bam\_arcs* track, connections between aligned mate pairs of a genomic read library show unusually large overlapping arcs that suggest an assembly inversion. The region of the suspected assembly inversion is highlighted using the highlight tool to the right of the navigation box, and the LSAA pull-down menu can be used to annotate the region as an inversion.

tation tools using JBrowse (3) and Apollo (25) for two *Bos taurus* genome assemblies, the reference genome assembly (UMD3.1.1) (26) and the alternate genome assembly (Btau.4.6.1) (24), as well as a data mining tool called BovineMine, based on the InterMine data warehousing platform (27). Since then we have been making continuous improvements to our existing tools, adding new tools and data and updating existing datasets.

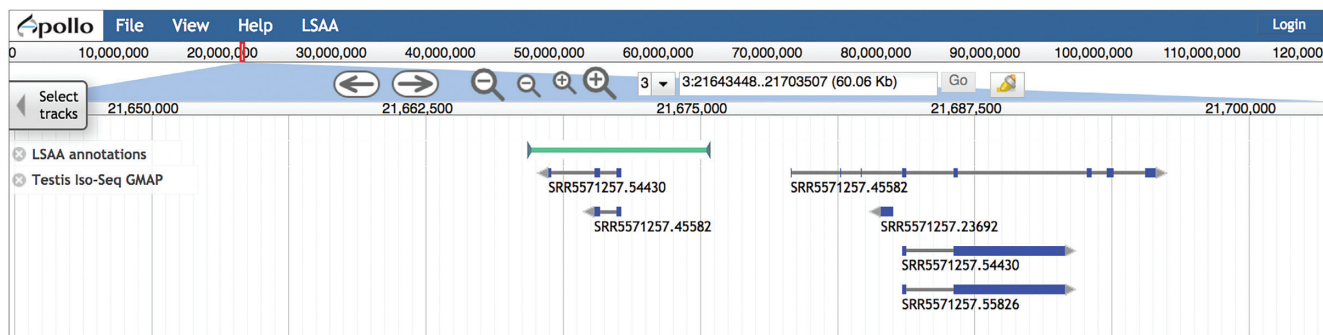
### LSAA: NEW TOOL FOR ANNOTATION OF GENOME ASSEMBLIES

Although sequencing technologies and genome assembly algorithms have improved over the years, reference genome assemblies still contain errors that can impact gene annotation (28). We have created the Locus-Specific Alternate Assembly (LSAA) tool, developed as a plugin for Apollo, which enables users to manually correct local assembly errors or provide data about genomic structural variants, such as breed specific variants, in the genome browser. The user-submitted LSAA annotations provide the locations and alternative sequences to correct assembly errors or describe variants. We added database objects for alternative loci to the Apollo 2.0 codebase, which utilizes the Chado schema (29), allowing LSAA to be saved in the Chado database. By leveraging Apollo, our tool allows users to curate LSAA while they annotate genes. LSAA user features include (i) a pull-down menu that opens a box in which users can enter coordinates and provide sequences and descriptions to create LSAA annotations (Figure 1); (ii) a Paired Arc viewer to assist the visualization of mate pair BAM files, with arcs

that connect the read pairs and can highlight possible assembly issues (Figure 1); (iii) a combined LSAA report and export tool that enables searching and filtering; (iv) a LSAA track with custom set of glyphs that represent different assembly changes (Figure 2).

An example of the application of the LSAA tool is provided in the Supplementary Data. Researchers often come across assembly errors in the process of manual gene annotation. As an example, an assembly inversion can be indicated when different segments of a single cDNA sequence align on opposite strands of the assembly or when the distance between mate pair reads in a genomic mate pair library align at larger or shorter distances from each other than expected (Figure 1). Supplementary Figure S1 shows a BLAST search with a cDNA sequence, resulting in an unexpected alignment pattern seen in JBrowse, followed by further investigation with additional JBrowse tracks. Supplementary Figures S2–S4 show the use of the LSAA tool to annotate an inversion, and Supplementary Figures S5–S6 show how to export and view submitted LSAA.

The LSAA tool provides visualizations to help users pinpoint assembly issues. Integration with a BLAST search tool enables visualizing BLAST high-scoring segment pairs (HSPs) on the browser (Supplementary Figure S1). A Paired Arc track displays genomic read alignments as arcs between read pairs, highlighting discordant read pairs as unusually large arcs (Figure 1 and Supplementary Figure S2). While the BLAST and Paired Arc visualizations are available prior to logging in to Apollo, users must be logged in to Apollo (described below) in order to submit and export LSAA annotations (Supplementary Figure S3). The



**Figure 2.** LSAA track in JBrowse. Once a tentative assembly inversion has been annotated, it is viewable as a track in JBrowse. The inversion is depicted as a green bar with inverted arrowheads.

menu for LSAA creation allows for the submission of locations for insertions, deletions, inversions and other changes, such as substitutions, as well as the entry of associated information, such as breed, individual, description and alternative sequence (Supplementary Figure S4). Once LSAA annotations are submitted, they are maintained, viewable, searchable, and downloadable but the reference assembly is not permanently changed. Results can be instantaneously shared with other bovine researchers. A filterable table of submitted LSAA annotations provides functions for navigating to LSAA regions in the browser; exporting LSAA FASTA sequences and schematic representations in JSON; and deleting LSAA (Supplementary Figure S5). LSAA annotations can be viewed on the browser by selecting the LSAA annotations track (Figure 2). Right clicking a LSAA feature opens a panel showing annotation details (Supplementary Figure S6). Code for the LSAA plugin for Apollo is available on GitHub (<https://github.com/elsiklab/Apollo-LSAA>) for use by other genome database providers.

### JBROWSE/APOLLO: NEW AND OLD ASSEMBLIES

Previously, BGD had genome browsers for the UMD3.1.1 bovine reference genome assembly and an alternate genome assembly, Btau.4.6.1. A new bovine reference genome, ARS-UCD1.2, with notably improved contiguity, was recently released (30). BGD currently provides JBrowse for both the new ARS-UCD1.2 reference and the former UMD3.1.1 reference genome, which can be selected from the JBrowse pulldown menu in the BGD navigation bar. The 'Alternate Assembly Alignment' tracks available in each browser show the results of whole genome alignment and chaining using BLAT and associated tools (31). Each browser includes tracks representing either the ungapped genome segments or the result of chaining the ungapped segments together. Clicking an ungapped segment opens a new JBrowse window of the alternate assembly, zoomed into the region of the aligned segment. Using this track in combination with viewing gene prediction tracks enables users to navigate across assemblies to examine genes in corresponding regions. The availability of browsers for both the UMD3.1.1 and ARS-UCD1.2 bovine reference genomes with large numbers of transcript and genomic data tracks will be valuable as researchers transition from the older to the newer bovine reference genome assembly.

Apollo annotation tools for the ARS-UCD1.2 genome assembly are available after logging in using a box in the upper right corner of the JBrowse window. The Apollo registration page, accessible from the Apollo pulldown menu on the BGD navigation bar, provides instructions for non-registered users who wish only to view existing annotations. Users must register using a valid email address to access all Apollo functions. With a user account, researchers can view and export all existing user annotations, manually edit genes and annotate new genes. An additional Apollo instance, the Bovine Apollo Demo, also available using the Apollo pulldown menu on the BGD navigation bar, provides full Apollo functionality with a guest login, but is intended for annotation practice, not for maintaining user submitted annotations.

### BGD BLAST: IMPROVED INTEGRATION WITH APOLLO

The BGD BLAST search tool, which is based on the Sequenceserver platform (6), is accessible from the BGD navigation bar. BLAST hits are linked to JBrowse based on match coordinates when a genome assembly is used as the search database (Supplementary Figure S1). Previously, the blast hits could be viewed only as highlighted regions in the genome. Now alignments of individual BLAST HSPs and their directionality can be viewed and used as evidence in a gene or LSAA annotation. Furthermore, when logged into Apollo, the HSP features can be dragged to the user annotation area to initiate gene annotations.

### GENE EXPRESSION DATA: UPDATED AND NEW

In the past, BGD had gene expression tracks in JBrowse and computed gene expression levels in BovineMine for UMD3.1.1, based on single-end RNA-seq data for 95 tissues (NCBI BioProject PRJNA263600). We have reprocessed the single-end RNA-seq data with the new reference genome, ARS-UCD1.2, and have added 40 paired-end RNA-seq datasets (NCBI BioProjects PRJNA379574 and PRJNA294306). Our methods for computing JBrowse tracks and BovineMine gene expression levels were updated from the previous TopHat/Cufflinks pipeline (32,33), to the improved Hisat2/StringTie pipeline (34,35). In addition, 23 long-read transcriptome datasets (Iso-Seq) were added



to JBrowse/Apollo (NCBI BioProjects PRJNA386670 and PRJNA434299). Multiple Iso-Seq and RNA-seq visualizations are available for each dataset.

## BOVINEMINE: ADDED FEATURES AND UPDATES

BovineMine has undergone five new releases since the previous BGD update publication (2). The current release is BovineMine v1.6. Each new release incorporated new and updated data, with better integration, improved queries and refined user interfaces. Detailed examples for the application of BovineMine were published (7) using the BovineMine v1.3 release, which is still available on the BGD homepage. BovineMine enhancements are described below.

### Multiple ruminant species and genome assemblies

BovineMine now hosts multiple genome assemblies of *Bos taurus* (UMD3.1.1 and ARS-UCD1.2), as well as genome assemblies of other ruminants: *Ovis aries* (Oar\_v3.1 and Oar\_v4.0) (36,37) and *Capra hircus* (ARS1) (38). Because the transition to a new genome can be difficult for researchers in the midst of ongoing projects, BovineMine has template queries that support conversion between gene sets and assemblies. Users have the choice of continuing research on the older assembly or using the new assembly. Template queries for converting between genomes and gene sets combined with the previously described List Tool (2) make it possible to perform meta-analysis of published research using both the old and new genome assemblies. The ability to convert between assemblies is especially important for the bovine research community, for which there is 10 years of published research on the older assembly.

With the availability of multiple assemblies per organism in BovineMine, we needed to modify the Regions Search tool, which allows users to perform a chromosome and coordinate based search for genomic features. In the past, the Regions Search tool allowed only one assembly per organism and users could select specific organism from a dropdown menu. The Regions Search tool now has an additional pulldown menu to select the assembly.

### New ortholog and metabolic datasets

BovineMine contains a new ortholog dataset called ORIS (ORthology Inference using Synteny), which includes bovine, human, mouse, sheep and pig genes (39). In addition, we incorporated a new metabolic reconstruction that was developed from the ORIS ortholog set based on the RECON human metabolic network reconstruction (40) using previously reported methods (41). BovineMine enrichment widgets allow testing gene lists for enrichment of metabolic subsystems and reactions.

### New ontologies

Previously the only ontologies supported by BovineMine were Sequence Ontology (42), Gene Ontology (43,44) and Mouse Anatomy Ontology (45). We have added several new ontologies, including Vertebrate Trait Ontology (46),

Evidence Code Ontology (47), Uberon Anatomy Ontology (48), Protein-Protein Interaction and Molecular Interaction Ontology (49), Livestock Breed Ontology (<https://www.animalgenome.org/bioinfo/projects/lbo/>) and Livestock Product Trait Ontology (<https://www.animalgenome.org/bioinfo/projects/lpt/>).

### Other BovineMine enhancements

In addition to the major BovineMine enhancements described above, several minor enhancements were implemented. There are multiple datasets where improvements were made in the data integration step. For example, QTL in BovineMine are now integrated with flanking SNP and relevant ontologies; expression metadata in BovineMine are now linked with relevant anatomy and tissue ontologies; orthology data in BovineMine can now be filtered by last common ancestor; and many new template queries have been added.

## MISCELLANEOUS BGD IMPROVEMENTS

We have made additional improvements and updates to the BGD website. The BGD website itself and all underlying applications have been moved to a new server, resulting in notable improvements in the responsiveness of BovineMine and JBrowse, and the BGD website has been completely revamped. Several software applications, including Apollo, InterMine and the Drupal content management system have been updated. The search tools previously accessible via 'Search and Annotation Tools' on the main BGD navigation bar, and designed to help annotators select tracks or chromosomal locations, have now been incorporated into BovineMine as template queries.

## ACCESSING BGD DATA AND TOOLS

BGD tools and data are freely available at <http://bovinegenome.org>, but some tools require login for full functionality. All data hosted at BGD, other than user submitted gene annotations, are accessible without logging in using JBrowse, BovineMine, Sequenceserver BLAST and the data download page. As described above, free Apollo registration is required for submitting or editing annotations using Apollo and the LSAA Tool. The Apollo registration page also provides guest login instructions for those who wish to view or export user annotations without editing. A separate Apollo instance, the Bovine Apollo Demo, is available with guest login for those who wish to practice annotating genes, but the practice annotations are not permanently maintained at BGD. BovineMine is freely available without logging in, but registering for a BovineMine account, by clicking 'Log in' on the BovineMine home page, provides several advantages. When users are logged in to BovineMine, query histories are automatically saved, and users can save queries, query templates and lists for later sessions and for sharing with other users. In summary, the only aspect of BGD that requires user registration is the ability to submit and modify annotations using the main (non-demo) bovine Apollo instance, whereas user registration for BovineMine is advantageous but not required.

## CITING BGD

Cite this article for the use of BGD tools such as BovineMine, JBrowse, Apollo, BLAST and code available on GitHub.

## CONCLUSIONS

To better serve the bovine research community, we have incorporated new and updated data, improved existing tools and developed new tools. In particular, the new LSAA Tool will allow the research community to curate both assembly errors and genomic structural variants among breeds and individuals, with submitted LSAA annotations immediately available for viewing and downloading by others. The LSAA annotations will both alert researchers of potential issues near genes of interest and provide valuable input to future bovine genome assembly upgrades. To support transition to the new reference genome, both BovineMine and JBrowse facilitate comparing old and new genomes and gene sets. Furthermore, BovineMine is the only resource that facilitates meta-analysis of published research results using the old and new bovine genome assemblies.

We will continue to enhance the Bovine Genome Database by adding new datasets and new tools to support bovine research community. BovineMine will be updated and released on an annual basis. Source code modifications will continue to be available on GitHub (<https://github.com/elsiklab/>). Most importantly, we encourage dataset suggestions and feedback from the bovine research community to guide future development.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

United States Department of Agriculture National Institute of Food and Agriculture [2013-67015-21202]; European Union's Seventh Framework Programme for research, technological development and demonstration [613689]. Funding for open access charge: University of Missouri.

*Conflict of interest statement.* None declared.

## REFERENCES

- Georges, M., Charlier, C. and Hayes, B. (2019) Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.*, **20**, 135–156.
- Elsik, C.G., Unni, D.R., Diesh, C.M., Tayal, A., Emery, M.L., Nguyen, H.N. and Hagen, D.E. (2016) Bovine genome database: New tools for gleaning function from the *Bos taurus* genome. *Nucleic Acids Res.*, **44**, D834–D839.
- Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
- Dunn, N.A., Unni, D.R., Diesh, C., Munoz-Torres, M., Harris, N.L., Yao, E., Rasche, H., Holmes, I.H., Elsik, C.G. and Lewis, S.E. (2019) Apollo: democratizing genome annotation. *PLoS Comput. Biol.*, **15**, e1006790.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Priyam, A., Woodcroft, B.J., Rai, V., Moghul, I., Mungala, A., Ter, F., Chowdhary, H., Pieniak, I.L., Gibbins, M.A., Moon, H. *et al.* (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases. *Mol. Biol. Evol.*, doi:10.1093/molbev/msz185.
- Hagen, D.E., Unni, D.R., Tayal, A., Burns, G.W. and Elsik, C.G. (2018) In: Kollmar, M. (ed). *Eukaryotic Genomic Databases: Methods and Protocols*. Springer, NY, pp. 211–249.
- Rajput, B., Pruitt, K.D. and Murphy, T.D. (2019) RefSeq curation and annotation of stop codon recoding in vertebrates. *Nucleic Acids Res.*, **47**, 594–606.
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S. *et al.* (2019) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
- Hu, Z.L., Park, C.A. and Reecy, J.M. (2019) Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.*, **47**, D701–D710.
- Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferson, T. *et al.* (2019) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **47**, D23–D28.
- UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I. *et al.* (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, **47**, D351–D360.
- Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R. and Birney, E. (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.*, **19**, 327–335.
- Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simao, F.A. and Zdobnov, E.M. (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.*, **47**, D807–D811.
- Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M. and Bateman, A. (2014) TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic Acids Res.*, **42**, D922–D925.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**, D590–D595.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B. *et al.* (2018) The Reactome Pathway Knowledgebase. *Nucleic Acids Res.*, **46**, D649–D655.
- Oughtred, R., Stark, C., Breitkreutz, B.J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R. *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
- Huntley, R.P., Sawford, T., Mutowo-Muullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J. and O'Donovan, C. (2015) The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.*, **43**, D1057–D1063.
- Nicolazzi, E.L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., Jorjani, H. *et al.* (2015) SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*, **16**, 283.
- Childers, C.P., Reese, J.T., Sundaram, J.P., Vile, D.C., Dickens, C.M., Childs, K.L., Salih, H., Bennett, A.K., Hagen, D.E., Adelson, D.L. *et al.* (2011) Bovine Genome Database: integrated tools for genome annotation and discovery. *Nucleic Acids Res.*, **39**, D830–D834.
- Bovine Genome Sequencing and Analysis Consortium, Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Elnitski, L. *et al.* (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, **324**, 522–528.
- Lee, E., Helt, G.A., Reese, J.T., Munoz-Torres, M.C., Childers, C.P., Buels, R.M., Stein, L., Holmes, I.H., Elsik, C.G. and Lewis, S.E. (2013)

- Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.*, **14**, R93.
26. 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. *et al.* (2009) A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.*, **10**, R42.
  27. Smith, R.N., Aleksic, J., Butano, D., Carr, A., Contrino, S., Hu, F., Lyne, M., Lyne, R., Kalderimis, A., Rutherford, K. *et al.* (2012) InterMine: A flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics*, **28**, 3163–3165.
  28. Florea, L., Souvorov, A., Kalbfleisch, T.S. and Salzberg, S.L. (2011) Genome assembly has a major impact on gene content: a comparison of annotation in two *Bos taurus* assemblies. *PLoS One*, **6**, e21400.
  29. Mungall, C.J., Emmert, D.B. and FlyBase, C. (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.
  30. Rosen, B., Bickhart, D., Schnabel, R., Koren, S., Elsik, C., Zimin, A., Dreischer, C., Schultheiss, S., Hall, R., Schroeder, S. *et al.* (2018) *Proceedings of the World Congress on Genetics Applied to Livestock Production, Vol. Molecular Genetics 3*. p. 802.
  31. Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
  32. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
  33. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
  34. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods.*, **12**, 357–360.
  35. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.*, **33**, 290–295.
  36. International Sheep Genomics Consortium, Archibald, A.L., Cockett, N.E., Dalrymple, B.P., Faraut, T., Kijas, J.W., Maddox, J.F., McEwan, J.C., Hutton Oddy, V., Raadsma, H.W. *et al.* (2010) The sheep genome reference sequence: a work in progress. *Anim. Genet.*, **41**, 449–453.
  37. Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., Wu, C., Muzny, D.M., Li, Y., Zhang, W. *et al.* (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, **344**, 1168–1173.
  38. 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.* (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.*, **49**, 643–650.
  39. Bekaert, M. and Conant, G.C. (2011) Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol. Biol. Evol.*, **28**, 1111–1121.
  40. Swainston, N., Smallbone, K., Hefzi, H., Dobson, P.D., Brewer, J., Hanscho, M., Zielinski, D.C., Ang, K.S., Gardiner, N.J., Gutierrez, J.M. *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, **12**, 109.
  41. Bekaert, M. and Conant, G.C. (2014) Gene duplication and phenotypic changes in the evolution of mammalian metabolic networks. *PLoS One*, **9**, e87115.
  42. Eilbeck, K., Lewis, S.E., Mungall, C.J., Yandell, M., Stein, L., Durbin, R. and Ashburner, M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.
  43. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
  44. The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
  45. Hayamizu, T.F., Baldock, R.A. and Ringwald, M. (2015) Mouse anatomy ontologies: enhancements and tools for exploring and integrating biomedical data. *Mamm. Genome*, **26**, 422–430.
  46. Park, C.A., Bello, S.M., Smith, C.L., Hu, Z.L., Munzenmaier, D.H., Nigam, R., Smith, J.R., Shimoyama, M., Eppig, J.T. and Reecy, J.M. (2013) The Vertebrate Trait Ontology: a controlled vocabulary for the annotation of trait data across species. *J. Biomed. Semantics*, **4**, 13.
  47. Chibucos, M.C., Mungall, C.J., Balakrishnan, R., Christie, K.R., Huntley, R.P., White, O., Blake, J.A., Lewis, S.E. and Giglio, M. (2014) Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database (Oxford)*, **2014**, bau075.
  48. Mungall, C.J., Torniai, C., Gkoutos, G.V., Lewis, S.E. and Haendel, M.A. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.*, **13**, R5.
  49. Krallinger, M., Vazquez, M., Leitner, F., Salgado, D., Chatr-Aryamontri, A., Winter, A., Peretto, L., Briganti, L., Licata, L., Iannuccelli, M. *et al.* (2011) The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. *BMC Bioinformatics*, **12**(Suppl. 8), S3.