

Bovine Genome Database: integrated tools for genome annotation and discovery

Christopher P. Childers¹, Justin T. Reese², Jaideep P. Sundaram¹, Donald C. Vile¹, C. Michael Dickens², Kevin L. Childs², Hanni Salih², Anna K. Bennett¹, Darren E. Hagen¹, David L. Adelson² and Christine G. Elsik^{2,*}

¹Department of Biology, Georgetown University, Washington, DC 20057 and ²Department of Animal Science, Texas A&M University, College Station, TX 77843, USA

Received August 16, 2010; Revised October 21, 2010; Accepted November 14, 2010

ABSTRACT

The Bovine Genome Database (BGD; <http://BovineGenome.org>) strives to improve annotation of the bovine genome and to integrate the genome sequence with other genomics data. BGD includes GBrowse genome browsers, the Apollo Annotation Editor, a quantitative trait loci (QTL) viewer, BLAST databases and gene pages. Genome browsers, available for both scaffold and chromosome coordinate systems, display the bovine Official Gene Set (OGS), RefSeq and Ensembl gene models, non-coding RNA, repeats, pseudogenes, single-nucleotide polymorphism, markers, QTL and alignments to complementary DNAs, ESTs and protein homologs. The Bovine QTL viewer is connected to the BGD Chromosome GBrowse, allowing for the identification of candidate genes underlying QTL. The Apollo Annotation Editor connects directly to the BGD Chado database to provide researchers with remote access to gene evidence in a graphical interface that allows editing and creating new gene models. Researchers may upload their annotations to the BGD server for review and integration into the subsequent release of the OGS. Gene pages display information for individual OGS gene models, including gene structure, transcript variants, functional descriptions, gene symbols, Gene Ontology terms, annotator comments and links to National Center for Biotechnology Information and

Ensembl. Each gene page is linked to a wiki page to allow input from the research community.

BACKGROUND

Cattle have provided nutrition to humans by converting plant material to muscle and milk for thousands of years. Furthermore, cattle have physiological characteristics with potential applications in biofuels and biomedical research. Cattle are ruminants, and thus achieve digestion of plant material by rumination. Food is fully digested in a four-compartment stomach including the rumen, a pregastric fermenter. Microbial organisms present in the rumen are being investigated as potential sources of enzymes that may be useful for the production of biofuels. For many years, the livestock industry has employed sire-based breeding systems and has performed routine measurements of economically important traits from very large numbers of animals, providing a genetic resource not available for most mammals. Many important traits such as weight gain, milk fat content and intramuscular fat (marbling) in cattle are quantitative traits (1,2). Several of the routinely measured livestock traits have relevance to human biology. For example, resistance to mastitis and parasites is relevant to human infectious disease resistance mechanisms. Susceptibility to bovine mastitis has been mapped to haplotypes that include the MHC DQ genes, which are involved in immune response (3,4). To decipher biological mechanisms underlying the quantitative trait loci (QTL), the presentation of genes and genetic markers in the same context as QTL is extremely important for candidate gene nomination.

*To whom correspondence should be addressed. Tel: +1 202 687 4485; Fax: +1 202 687 4662; Email: ce75@georgetown.edu
Present addresses:

Justin T. Reese and Christine G. Elsik, Department of Biology, Georgetown University, Washington, DC 20057, USA.

Kevin L. Childs, Department of Plant Biology, Michigan State University, 166 Plant Biology Building, East Lansing, MI 48824, USA.

David L. Adelson, School of Molecular and Biomedical Science, School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, South Australia 5005, Australia.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

The 7.1X genome assembly of a Hereford cow produced by the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) has now allowed years of work in cattle QTL mapping to be associated with genes and other genomic features (5).

The initial goal of the Bovine Genome Database (BGD) project was to support the Bovine Genome Sequencing and Analysis Consortium by providing web-based tools for annotation of the bovine genome, and a portal to collect and organize the annotations. In addition, we have provided a web page for posting and obtaining consortium data sets. We are still working to improve gene annotations and to incorporate additional applications to integrate new data types. Here, we present usage examples for annotation and discovery of bovine genes. The BGD annotation system may be used in the classroom to teach manual annotation methods or fundamental aspects of eukaryotic gene structure.

DATABASE CONTENT

BGD currently provides access to bovine genome assembly Btau_4.0 (5,6) and the bovine Official Gene Set version 2 (OGSv2). OGSv2 is composed of 23 633 gene models, 3164 of which were manually annotated. In addition to gene models and associated functional annotations, BGD includes protein homolog alignments, complementary DNA (cDNA) alignments, DNA repeats, single-nucleotide polymorphisms (SNPs), microsatellite markers and QTL intervals. Descriptive information is associated with OGSv2 genes whenever possible. To facilitate computational annotation, we produced a mapping of OGSv2 identifiers to National Center for Biotechnology Information (NCBI) and Ensembl identifiers for overlapping gene loci, with criteria for the locus in question that (i) Ensembl or NCBI RefSeq coding sequence coordinates must overlap OGSv2 coding sequence coordinates and (ii) the relationship between the OGSv2 and Ensembl or RefSeq gene locus must not include a split or merged gene model. The mapping file allows BGD gene pages to link to NCBI and Ensembl. In addition, NCBI links to BGD gene pages. GO annotations and gene symbols are obtained automatically on a weekly basis from NCBI for OGSv2 genes mapped to NCBI genes. For OGSv2 genes that do not overlap NCBI genes, GO annotations are obtained for genes that overlap Ensembl genes using EnsMart (7). Fasta sequence files for the manually annotated subset of OGSv2 and for the entire OGSv2 set, as well as an identifier mapping file, are available for download on the BGD website. The bovine OGS is updated following the release of new bovine assemblies and gene sets at NCBI and Ensembl. In the future, updates will occur annually.

BGD USE CASES

In the following sections, we provide examples to illustrate the use of BGD for gene annotation and for mining the bovine genome. Figures are provided in supplementary data.

Annotating bovine genes

Setting up: registering on BGD and configuring Apollo. Before beginning the annotation process, new users need to register and configure their computer for the annotation software. Users should register as an annotator on the community annotation site (http://bovinegenome.org/?q=annotator_login). By registering, users will have access to submit annotations, query and view the submitted annotations, view the list of priority genes and modify annotations submitted by that user. BGD annotators use the Apollo Annotation Editor (8) to create and submit annotations. Apollo is a java-based annotation editor that was originally developed for Flybase (9) and later was incorporated into the Generic Model Organism Database (GMOD) project. Apollo needs to be installed and set up for use with BGD before it may be used with BGD data. First, the user must download the latest version of Apollo (<http://apollo.berkeleybop.org/>). The next step is to download the configuration files that Apollo needs to be able to connect to the BGD database. The configuration files, along with additional instructions, may be found at http://bovinegenome.org/?q=apollo_files. After Apollo has been configured, it can directly access the BGD annotation database, including the reference genome sequence and all the available evidence.

Using BGD BLAST as a starting point for annotation. One of the most common ways to initiate an annotation project is to start with a sequence of interest such as a cDNA or expressed sequence tag (EST), or protein homolog. The BGD Basic Local Alignment Search Tool (BLAST) (10) web page can be accessed by selecting 'BLAST' in the 'Tools and Resources' pull-down menu. BGD BLAST databases include different bovine assemblies, different gene prediction sets and the bovine OGS. Assembly Btau_4.0 and OGSv2 are assembly and OGS that are currently supported, although BGD still offers assembly Btau_3.1 and OGSv1 as legacy BLAST databases. To determine if the gene of interest is in the OGS, the user would search the Bovine OGSv2 protein or cDNA BLAST database. BGD BLAST output has been customized to provide direct links to the gene pages for OGSv2 genes matching the query. If the gene model needs to be manually revised, the user may find sequence coordinate information by searching the BGD gene pages using the OGSv2 ID. Alternatively, the location on the genome of a sequence of interest may be mapped using BLAST to search either the 'Btau_4.0 Scaffolds' or 'Btau_4.0 Assembled Chromosomes' BLAST databases. BGD BLAST output to the sequence assemblies has been customized to contain links to the NCBI record for the scaffold or chromosome record. Output from genome BLAST database alignments also contain links to view the region in GBrowse (11), including the aligned sequence as an additional data track. The GBrowse view contains all of the available evidence, allowing users to rapidly examine how similar the queried sequence aligns to the current evidence. Tracks for the BLAST hits on the genome browsers are maintained using cookies, so that the

user may view multiple BLAST hits on the browser at the same time, and/or review the hits in a later session.

Accessing BGD GBrowse resources. In addition to connecting from BLAST output, GBrowse may be directly accessed from the BGD home page, under the 'Genome Browsers' pull-down menu. Both scaffold and chromosome coordinate systems are currently supported for assembly Btau_4.0. Users may enter a scaffold (or chromosome) ID into the search box to view the data for any region of the genome assembly. If a feature name (such as OGSv2 ID) is known, it can also be directly queried using the same search box. Although assemblies older than Btau_4.0 are no longer actively supported, GBrowse is still available for assembly 3.1 chromosome and scaffold coordinate systems.

Using BGD annotation utilities. BGD has additional tools (http://bovinegenome.org/?q=annotation_tools) to assist researchers in integrating information from multiple sites [e.g. NCBI (12), Ensembl (13), UCSC Genome Browser (14), BGD] despite differences in identifiers for sequence assembly components and different assembly coordinate systems (scaffolds versus chromosomes). For example, BGD uses identifiers provided by BCM-HGSC for scaffolds and chromosomes, because whenever possible these identifiers indicate the chromosome and the scaffold order along a chromosome. NCBI assigns assembly sequence accessions that fit within their naming conventions. In addition, different resources provide different assembly components with different coordinate systems in their genome browsers. BGD Apollo uses scaffolds to reduce the memory required on users' computers. Ensembl and the UCSC Genome Browser use whole chromosome assemblies. UCSC creates a single long pseudo-chromosome of concatenated unassigned scaffolds, while BGD maintains separate unassigned scaffolds. BGD provides tools for converting between all of these systems. To illustrate, say an annotator finds evidence for an interesting gene on chromosome NC_007299 between bases 44 500 and 66 000 at NCBI and decides to annotate it. The user would enter the NCBI accession and coordinates in the 'Chromosome to Scaffold Conversion Tool' to determine that the region corresponds to Chr1.1, bases 44 500–66 000. The annotation utilities page also contains tools for retrieving the length of a scaffold, which helps users avoid loading more sequence than exists on a scaffold. There is also a tool for looking up protein homolog annotations, multiple flash tutorials that illustrate how to use Apollo to annotate genes and links to external databases and tools.

Modifying or creating gene models using BGD Apollo. Once the genomic region has been determined, the annotation process continues with the Apollo Annotation Editor. After starting Apollo, the user changes the data source to 'Chado Database' and the database name to 'Bovine Genome Assembly 4'. At this point there are two ways to proceed. The first way is to enter the scaffold name and coordinates for the region in which the gene resides. Although the Apollo menu labels

this as 'chromosome', BGD uses the scaffold coordinate system for Apollo annotation. The second way to continue is to change the 'Type of Region' to 'gene', then type in the OGSv2 ID to load the region of the scaffold in which the gene resides, and an additional 50 000 bases of sequence on each side of the gene. The flanking sequence is useful if the gene model must be extended by adjusting the coding sequence to reflect an alternative start/stop site, merging two gene models or adding untranslated regions (UTRs).

Once Apollo finishes loading the desired region, all of the available evidence becomes visible, including predicted gene models from NCBI RefSeq (15), Ensembl (16), four different prediction tools [Fgenesh, Fgenesh++ (17,18), GENEID (19) or SGP2 (20)] and OGSv2 gene models. Evidence tracks also include alignments of ESTs and cDNAs produced using Genomic Mapping and Alignment Program (GMAP) (21) and alignments of protein homologs produced using Exonerate (22). GMAP and Exonerate are both splice modeling alignment programs, so the tracks they produce can aid in the identification of splice sites. The user determines whether a gene model should be annotated based on the available evidence, and drags the evidence track most representative of the final gene model into the blue 'working area'. The user then checks and adjusts exon boundaries and untranslated regions (UTR) if necessary. The user may add additional information, including homolog IDs, gene symbols or synonyms, by right clicking on the gene model and clicking the 'Annotation info editor' option. The user may provide specific comments about the gene including reasons for adjusting the gene model or comments regarding ambiguities for review by the BGD annotation curator. We have preloaded commonly used comments so the user can select these from a dropdown menu. Using these 'canned' comments promotes the use of standardized language, which simplifies automated processing. After the annotation is complete, the user saves it using a pull-down menu. Selecting 'Chado Database' as the format and 'Bovine Genome Assembly 4' as the database causes to be loaded directly into the BGD Chado database. The user may also save the annotation locally as a Chado-XML file.

Discovery

Identifying genes that underlie QTL. The Bovine QTL viewer (23,24), containing QTL data curated from literature, allows users to search for traits on one or more chromosomes, or to directly search for a specific QTL. In order to access the QTL viewer from the BGD home page, the user clicks on the 'Tools and Resources' tab then click 'Bovine QTL Viewer', and is then redirected users to the QTL Viewer Site. The users login as 'guest' to reach the main interface for searching the QTL database. If the user does not know the exact terminology for the trait of interest, they can select one or more categories and any number of chromosomes to search through. QTL Viewer displays the results of the search as highlighted regions along the chromosomes. Clicking on a QTL region on the main map opens a

zoomed view showing specific QTLs, along with their names. Clicking on a QTL name opens the main page for that QTL. The QTL page displays information on the position, statistical significance, references, specific markers that lie within the QTL, other data relevant to the locus and links to the BGD GBrowse. For example, if the user is interested in traits related to dairy production, they could select 'Milk Fat', 'Milk Protein' and 'Milk Yield'. Chromosomes can also be selected. Without prior expectations for where such traits might lie, the user would select 'All Chromosomes'. Using the zoom feature in the resulting chromosome display, would see that Chromosome 27 has two short QTLs for Milk Yield and Milk Fat that overlap, along with a larger QTL for Milk Protein. After clicking one of the QTLs the user would see that there are small overlapping QTLs for all three categories and a larger QTL for Milk Protein. Clicking the larger Milk Protein QTL opens the record for that particular locus. From here the user can click 'Assembly 4.0 View' to open the BGD Chromosome GBrowse. Once in GBrowse the user can view different tracks to identify genome features that underlie this QTL. After zooming in, the user would many spliced EST alignments and several gene models within the locus, providing a starting point for further research.

Searching for information on Gene Pages. The Bovine OGSv2 contains 23 632 gene models. It consists of gene models from Ensembl (16), RefSeq (15), GLEAN (25) consensus gene set, gene models produced using full-length cDNA alignments and submitted manual annotations. Details on the creation of OGSv1 and OGSv2 are provided in (5) and (26), respectively. The gene pages are a new Ruby on Rails application developed in our laboratory. They include a full mapping of the Chado Database and a robust search function allowing users to approach the OGSv2 data from many directions. The initial page simply contains a search box and some example search terms and methods to help users get started.

For example, if the user is interested in genes that have been associated with dephosphorylation, they can simply enter the term 'dephosphorylation' and click Search. This search results in 84 hits, some of which are not directly related to the term in which we are interested. Alternatively, the user may first check the Amigo Gene Ontology (27) site to retrieve the GO ID to identify the GO ID for dephosphorylation (GO:0016311) and then use the GO ID in the gene search box to retrieve a smaller, and likely more relevant, set of 13 genes. If the OGSv2 gene symbol is already known, entering it into the search box will take the searcher straight to the relevant gene page. BGD obtains gene symbols for OGSv2 automatically from NCBI on a weekly basis for genes that have a clear one to one orthology relationship with human genes. Users can also search for OGSv2 genes using submitter name, user ID or the provisional ID used during the community annotation phase before the final OGSv2 IDs were created.

Each gene page is made up of several sections: an overview of the gene including gene symbols, overlapping NCBI or Ensembl locus (or gene) identifier, genomic coordinates, GO annotations and a dynamically generated image showing the gene models. Clicking the 'Show Evidence' button above the GBrowse gene model image displays additional evidence that overlap the gene model. Clicking the View in GBrowse button opens the region containing the gene in the BGD assembly 4.0 Scaffold GBrowse. The second section of the gene page contains information for the alternative transcripts. Each transcript is displayed, along with information specifically associated with that variant, including the source for that annotation, such as 'Manual' or 'GLEAN', and any additional information for that transcript, such as exon coordinates. The cDNA and translated sequences are available for each transcript in plain-text, HTML or XML formats. The last section of each gene page is a collection of all the transcript and protein sequences, for convenient access. Finally, each gene page provides a link to the wiki page for that gene. The wiki allows researchers to take an active part in contributing knowledge to the database. Users must register to use the wiki before they can add comments about a gene, even if they have already registered for annotation.

CONTENT MANAGEMENT

One of the most frequent users of BGD is our curator. In addition to using Apollo to review submitted genes, the curator updates the site with news and literature.

BGD is implemented using the Drupal Content Management System (<http://drupal.org>), which greatly simplifies creating, updating and maintaining information on the website, so experience in Web programming is not required of the curator. Page creation is handled through a simple web-based form with standard input fields. Although no experience with HTML code is needed for page creation, Drupal does support the use of HTML tags in the page body. Drupal has a very fine-grained permission system, so roles may be created and given very specific permissions to be able to access or edit only certain parts of the site. For example, BGD has roles defined for site administrators and also for site contributors. Contributors have access to create and edit content, while administrators also have additional privileges, such as to modify the sites appearance. The inclusion of different standard modules allows for the different features available at BGD, like a custom content type called 'News', which only displays in the News section of the front page, and is set to display new content at the top, pushing down older entries. Only the first several news items are available on the front page. Site themes are easier to create and adjust, as all of the content is stored in a database instead of in static pages.

FUTURE DIRECTIONS

The Bovine Genome Database project is ongoing. We will continue to improve the annotation of the bovine genome and will incorporate the next assembly from BCM-HGSC (Btau_4.2). We will also annotate and create a genome browser for an alternative assembly, UMD3.1 (28). This will require updating the OGS and mapping manual annotations to the new assemblies. We will incorporate RNA-Seq data into the development of the new OGS. As genomes for additional livestock species continue to become available, we will deploy tools for comparison, such as synteny viewers. Finally, we plan to create tools for mining SNP and haplotype data, including a browser and a Biomart-based data mining tool (29).

AVAILABILITY

BGD is publicly accessible at <http://BovineGenome.org>. Using annotation tools and submitting comments to the wiki require registration.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

The United States Department of Agriculture National Institute of Food and Agriculture (2007-35616-17882 to C.G.E. and D.L.A.); the Kleberg Foundation; Texas AgriLife; start-up funds from Georgetown University. Funding for open access charge: USDA National Institute of Food and Agriculture (2007-35616-17882).

Conflict of interest statement. None declared.

REFERENCES

- Ashwell,M.S., Heyen,D.W., Sonstegard,T.S., Van Tassell,C.P., Da,Y., VanRaden,P.M., Ron,M., Weller,J.I. and Lewin,H.A. (2004) Detection of quantitative trait loci affecting milk production, health, and reproductive traits in holstein cattle. *J. Dairy Sci.*, **87**, 468–475.
- Keele,J.W., Shackelford,S.D., Kappes,S.M., Koohmaraie,M. and Stone,R.T. (1999) A region on bovine chromosome 15 influences beef tenderness in steers. *J. Anim. Sci.*, **77**, 1364–1371.
- Mallard,B.A., Dekkers,J.C., Ireland,M.J., Leslie,K.E., Sharif,S., Vankampen,C.L., Wagter,L. and Wilkie,B.N. (1998) Alteration in immune responsiveness during the peripartum period and its ramification on dairy cow and calf health. *J. Dairy Sci.*, **81**, 585–595.
- Park,Y.H., Joo,Y.S., Park,J.Y., Moon,J.S., Kim,S.H., Kwon,N.H., Ahn,J.S., Davis,W.C. and Davies,C.J. (2004) Characterization of lymphocyte subpopulations and major histocompatibility complex haplotypes of mastitis-resistant and susceptible cows. *J. Vet. Sci.*, **5**, 29–39.
- Bovine Genome Sequencing and Analysis Consortium. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, **324**, 522–528.
- Liu,Y., Qin,X., Song,X.Z., Jiang,H., Shen,Y., Durbin,K.J., Lien,S., Kent,M.P., Sodeland,M., Ren,Y. *et al.* (2009) Bos taurus genome assembly. *BMC Genomics*, **10**, 180.
- Kasprzyk,A., Keefe,D., Smedley,D., London,D., Spooner,W., Melsopp,C., Hammond,M., Rocca-Serra,P., Cox,T. and Birney,E. (2004) EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.*, **14**, 160–169.
- Lewis,S.E., Searle,S.M., Harris,N., Gibson,M., Lyer,V., Richter,J., Wiel,C., Bayraktaroglu,L., Birney,E., Crosby,M.A. *et al.* (2002) Apollo: a sequence annotation editor. *Genome Biol.*, **3**, RESEARCH0082.
- Tweedie,S., Ashburner,M., Falls,K., Leyland,P., McQuilton,P., Marygold,S., Millburn,G., Osumi-Sutherland,D., Schroeder,A., Seal,R. *et al.* (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res.*, **37**, D555–D559.
- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
- Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetverin,V., Church,D.M., Dicuccio,M., Federhen,S. *et al.* (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **38**, D5–D16.
- Flicek,P., Aken,B.L., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S. *et al.* (2010) Ensembl's 10th year. *Nucleic Acids Res.*, **38**, D557–D562.
- Rhead,B., Karolchik,D., Kuhn,R.M., Hinrichs,A.S., Zweig,A.S., Fujita,P.A., Diekhans,M., Smith,K.E., Rosenbloom,K.R., Raney,B.J. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.*, **38**, D613–D619.
- Pruitt,K.D., Tatusova,T., Klimke,W. and Maglott,D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
- Curwen,V., Eyraes,E., Andrews,T.D., Clarke,L., Mongin,E., Searle,S.M. and Clamp,M. (2004) The Ensembl automatic gene annotation system. *Genome Res.*, **14**, 942–950.
- Salamov,A.A. and Solovyev,V.V. (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Res.*, **10**, 516–522.
- Solovyev,V. (2007) In Balding,D.J., Bishop,M. and Cannings,C. (eds), *Handbook of Statistical Genetics*. John Wiley & Sons, Chichester, UK, Hoboken, NJ, pp. 97–159.
- Blanco,E., Parra,G. and Guigo,R. (2007) Using geneid to identify genes. *Curr. Protoc. Bioinformatics*, Chapter 4, Unit 4.3.
- Parra,G., Agarwal,P., Abril,J.F., Wiehe,T., Fickett,J.W. and Guigo,R. (2003) Comparative gene prediction in human and mouse. *Genome Res.*, **13**, 108–117.
- Wu,T.D. and Watanabe,C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, **21**, 1859–1875.
- Slater,G.S. and Birney,E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.
- Polineni,P., Aragona,P., Xavier,S.R., Furuta,R. and Adelson,D.L. (2006) The bovine QTL viewer: a web accessible database of bovine Quantitative Trait Loci. *BMC Bioinformatics*, **7**, 283.
- Salih,H. and Adelson,D.L. (2009) QTL global meta-analysis: are trait determining genes clustered? *BMC Genomics*, **10**, 184.
- Elsik,C.G., Mackey,A.J., Reese,J.T., Milshina,N.V., Roos,D.S. and Weinstock,G.M. (2007) Creating a honey bee consensus gene set. *Genome Biol.*, **8**, R13.
- Reese,J.T., Childers,C.P., Sundaram,J.P., Vile,D.C., Dickens,C.M., Childs,K.L. and Elsik,C.G. (2010) Bovine Genome Database: supporting community annotation and analysis of the Bos taurus genome. *BMC Genomics*, **11**, 645.
- Carbon,S., Ireland,A., Mungall,C.J., Shu,S., Marshall,B. and Lewis,S. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.
- 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.
- Haider,S., Ballester,B., Smedley,D., Zhang,J., Rice,P. and Kasprzyk,A. (2009) BioMart Central Portal—unified access to biological data. *Nucleic Acids Res.*, **37**(Web Server issue), W23–W27.