We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Using Bacterial Artificial Chromosomes to Refine Genome Assemblies and to Build Virtual Genomes

Abhirami Ratnakumar[1,2], Wesley Barris[1,3],
Sean McWilliam[1] and Brian P. Dalrymple[1]
[1]CSIRO Livestock Industries, St. Lucia, QLD
[2]now at the Department of Medical Biochemistry and Microbiology,
Uppsala University, Uppsala
[3]now at Cobb-Vantress, Siloam Springs, Arkansas
[1]Australia
[2]Sweden
[3]USA

1. Introduction

Recent years have seen an explosion in the sequencing of genomes, including those of ruminants. A number of assemblies of the sequence of the bovine genome are now available (Elsik, et al., 2009; Zimin, et al., 2009). Although the sheep genome sequence is not such a high priority, the International Sheep Genomics Consortium (ISGC_website) has a long term strategy to develop a number of tools for the application of genomics in sheep research and breeding (Archibald, et al., 2010). We have demonstrated recently how comparative genomics and Bacterial Artificial Chromosome (BAC)-libraries can be used to construct detailed virtual genomes as a framework for genome assemblies of related species (Dalrymple, et al., 2007). As new and improved genome assemblies of the genomes contributing to an initial virtual genome assembly are produced, the virtual genomes will need to be regularly updated to incorporate the latest available information. In the original analysis, three genomes (bovine, dog and human) with various levels of coverage and stages of assembly were used (Dalrymple, et al., 2007). With the availability of increasing numbers of assemblies, the benefit of using more than three genomes, or the most appropriate evolutionary distances of the genomes, is not immediately clear. Here we describe the construction of a modified version of the bovine Btau3.1 assembly using cattle and sheep BACs and the use of this assembly in the construction of an updated virtual sheep genome, combining information from the original sheep virtual genome (vsg 1.2) and the horse (Wade, et al., 2009) and dog (Lindblad-Toh, et al., 2005) genomes. The impact of inclusion of additional genome sequences is analysed. The approach described here for sheep is an example of an approach which can be applied more broadly to genomes of any source, for example for the fish species, tilapia (Soler, et al., 2010) and catfish (Liu, et al., 2009). Indeed, the same principles also apply to the detection of differences between different individuals of the same species.

2. Materials and methods

2.1 Data sources and sequence search parameters

All genome sequences, except Btau3.#x versions, were downloaded from the UCSC comparative genomics website (UCSC; Fujita, et al., 2011). The full set of BAC-end sequences (BESs) from the CHORI-243 sheep BAC library, deposited in GenBank with the following accession numbers; CL632218-CL639051, CZ920079-CZ926973 and DU169919-DU532729 (Dalrymple, et al., 2007), were filtered to remove duplicate sequences and to identify the set of high confidence BACs (Ratnakumar, et al., 2010a). The filtered set of sheep BAC-end sequences were aligned to the lower case masked versions of the bovine genome assembly (Btau3.1) and the revised bovine genome assembly (Btau3.5x) using MegaBLASTn with the following optimised parameters: -r 1 -q -1 -X 40 -W 8, as previously described (Ratnakumar, et al., 2010b). The filtered set of sheep BESs were aligned to the lower case masked versions of the dog genome sequence assembly (canFam2), the horse genome sequence assembly (equCab1) and to the human genome sequence assembly (hg17) using BLASTn with the following parameters: -W 7 -r 17 -q -21 -G 29 -E 22 -X 240 -e 1 -f 280 -F m -U T and -z 3076781887 (human) and -z 2531657226 (dog), as previously described (Dalrymple, et al., 2007). No cut offs were applied to the BLAST output except that for each sheep BAC-end sequence only the best hit from each of the genomes was used for the next steps. If two BESs hits to the same genome assembly with equal scores were obtained the hit on the same chromosome as the best hit for the BES determined from the other end of the BAC was retained. If more than two hits with equal scores were obtained the BES hit was discarded.

The BESs from the CHORI-240 cattle BAC library, GenBank accession numbers; BZ830806-BZ891831, BZ896446-BZ956676, CC447354-CC447937, CC466118-CC470858, CC470880-CC596504, CC761663-CC775995, CG917936-CG918393, CG976420-CG992944, CL603252-CL610093, CW848133-CW848163, CZ012846-CZ027312 (Snelling, et al., 2007), were aligned to the cattle and virtual sheep genome sequences using BLAT (Kent, 2002).

Bovine genome assembly Btau3.1 sequence contigs were aligned to the human, dog and horse genomes using MegaBLASTn as described above.

2.2 Genome coordinate conversion

The coordinates from the mapping of the sheep BESs to the dog and human genomes were converted to the framework of the bovine genome assembly Btau3.1 using the LiftOver utility (LiftOver; Fujita, et al., 2011) and the canFam2 to Btau3.1 and hg17 to Btau3.1 coordinate conversion chain files respectively, also downloaded from UCSC genome bioinformatics site (UCSC; Fujita, et al., 2011). If the initial application of LiftOver was not successful for a region of the genome, regions of 100 bases either side of the BAC-end sequence were taken and positioned using LiftOver (pseudoliftOver). If this was again unsuccessful the process was repeated in steps of 100 bases until a successful application of the LiftOver utility for a region was achieved, or a distance of 10kb was reached (Dalrymple, et al., 2007).

Coordinate conversion (chain) files able to be read by the LiftOver utility to convert bovine genome assembly Btau3.1 coordinates to bovine genome assembly Btau3.#x version coordinates were built based on the revised order of Btau3.1 contigs and scaffolds in Btau3.#x version. Similarly a coordinate conversion file to convert Btau3.5x coordinates to virtual sheep genome assembly coordinates was built based on the order of Btau3.5x scaffolds in the virtual sheep genome.

2.3 Assigning BACs to groups and building BAC contigs

BACs were assigned to the groups; "tail-to-tail", "tail-to-head" etc. on the basis of the relative orientations of the two BESs from each BAC on the relevant genome assembly and the distance apart of the BESs. "Outsize" BACs were those with the two BESs mapped to the same chromosome in the relevant genome assembly and less than 10 kb, or more than 200 kb, apart. Data processing was undertaken using a series of Perl scripts. BACs with both BESs mapped to the genome, but mapped to two different chromosomes, were assigned to the "breaks" group. BACs with only one BES mapped to the genome were assigned to the "unpaired" group.

BAC-comparative genomic contigs (BAC-CGCs) were constructed for the BACs from each species mapped to each genome assembly using Perl scripts to process the data (Dalrymple, et al., 2007). Starting from the beginning of each chromosome the first BAC that overlapped with a second BAC was identified, the BAC-CGC was extended until no further overlapping BACs were identified. This process was repeated along the chromosome until the last BAC mapped on the chromosome was reached. The process was repeated for each chromosome in the genome assembly.

2.4 Construction of Btau3.5x

Using Perl scripts and the data set of the mapping of the bovine BESs to the scaffolds of the Btau3.1 genome assembly an initial minimization of the number of non-tail-to-tail BACs was undertaken. The scripts started with the first scaffold on chromosome 1 of the assembly and by testing the number of BAC links between this scaffold and all other scaffolds in the assembly identified the most likely adjacent scaffold and the orientation of the scaffold based on maximising the number of tail-to-tail BACs. Two or more linking tail-to-tail BACs without overlapping BES mapping coordinates on both scaffolds were required to continue the chain. Only high confidence bovine BACs (Ratnakumar, et al., 2009) were used in the assembly. Adjacent scaffolds assigned to the same chromosome in the Btau3.1 assembly were preferred over a more highly linked scaffold assigned to another chromosome, if the preferred scaffold on the original chromosome was itself linked to an adjacent scaffold on the original chromosome. If no scaffold assigned to the same chromosome as the rest of the chain was linked into the chain by BACs, or the less strongly linked scaffold from the same chromosome terminated the chain, the most highly linked scaffold from another chromosome was incorporated. If the newly added scaffold was linked back to the original chromosome at the next step of scaffold incorporation it was retained in the chain, otherwise the chain was terminated and the chromosome changing scaffold was also removed from the scaffold chain.

For each scaffold in the chain BAC-links from both ends of the scaffold were assessed to enable to inclusion of scaffolds preceding the initiating scaffold, or located between two scaffolds in a chain, but which were only linked to an adjacent following scaffold. The scaffold chain building process was continued until it was terminated with a scaffold not linked by two or more BACs to another scaffold. The penultimate scaffold in the chain was then tested for BAC links to a second scaffold and incorporated if it met the criteria described above. The chain building process was then continued. If no second linked scaffold could be identified the scaffold chain building was terminated. The next unincorporated scaffold from the same chromosome of the Btau3.1 assembly was then used to initiate the next scaffold chain. When all scaffolds from the first chromosome had been tested the first scaffold from the next chromosome was used and the process repeated until all scaffolds assigned to a chromosome of the Btau3.1 assembly had been tested.

Once the scaffold chain assembly had been completed the scaffolds not assigned to chromosomes in the Btau3.1 assembly (the UnChr) were then linked into the scaffold chains in a similar, but separate process. The resulting scaffold chains were then ordered and oriented using the consensus of the mapping of the order of the BACs in the physical bovine BAC map (Snelling, et al., 2007) to the BACs in the bovine scaffold chains. The initial data set Btau3.1x was then displayed as a browseable genome using Gbrowse (Stein, et al., 2002) to allow the integrity of the assembly of the scaffolds to be visually assessed. Genome contigs, scaffolds, bovine BAC mapping positions were displayed as separate groups. Clusters on non-congruent BACs (i.e. not tail-to-tail) identified regions with remaining assembly problems.

Using Perl scripts and the data set of the mapping of the sheep BESs to the Btau3.1 genome assembly, including BES mapping data integrated onto the Btau3.1 assembly from the horse, dog and vsg1.2 assemblies sheep BACs were assigned to tail-to-tail etc. groups and displayed on the Btau3.1x genome browser in a series of tracks. Positions of the BESs mapped to the separate genomes were integrated on Btau3.1 as previously described (Dalrymple, et al., 2007).

The mappings of the bovine genome assembly Btau3.1 sequence contigs to the human, dog and horse genomes were displayed as separate tracks on the Btau3.1x genome browser using the UCSC chromosome colour scheme (Fujita, et al., 2011) to identify the chromosome of best match in the relevant species. Asymmetric symbols were used to represent the orientation of the mapping of the contigs to the human, dog and horse genomes relative to the bovine genome. The chromosomal coordinates of the mapping in the non-bovine genome were also readily accessible to the users of the browser using mouse-over and mouse-click display boxes. This information was used in the manual refinement of the assembly, in particular in the definition of scaffold split points for the insertion of other scaffolds and/or the inversion of small numbers of adjacent contigs within a scaffold, where extensive use was made of comparative genomics information at the level of the sequence contigs.

Subsequently four major rounds of revision and refinement of the bovine genome assembly were undertaken manually and decisions on the chromosomal assignment, order of scaffolds and orientation of scaffolds and of sequence contigs were made based on the cattle and sheep BAC mapping and the comparative genomics. Generally in cases of ambiguity parsimony was applied. For the construction of each new version of the assembly changes were recorded in an Excel spreadsheet and Perl scripts were used to convert the Excel spreadsheet into a genome assembly agp file (AGP_file_specification). The agp file was used to generate the sequence of the genome assembly, the coordinate conversion chain file (for use by the LiftOver utility) and the contig and scaffold tracks for the genome browser version for the new assembly. For each successive version of the revised assembly of the bovine genome the manual revision was undertaken interactively using the tracks on the genome browser to make decisions.

2.5 Construction of the virtual sheep genome vsg2.0

To generate the virtual sheep genome assembly the mid point between each pair of BAC-CGCs built using sheep BACs on the bovine Btau3.5x genome assembly was identified. If the mid point was located in a gene (NCBI human RefSeq mRNAs (NCBI_RefSeq) were used to define the extent of a gene) the position closest to the midpoint and not in a gene was identified. The flanking BAC-CGCs were then extended to this point, or in the case of the first and last BAC-CGCs on a chromosome to the start or end coordinate of the

chromosome. Thus all nucleotides in the bovine genome sequence were included in a block and therefore the virtual sheep genome sequence is exactly the same length as the bovine Btau3.5x genome sequence.

The order and orientation of the bovine genome assembly Btau3.5x-based sheep BAC-CGCs in the vsg2.0 was determined on the location and organisation of the sheep linkage map markers (Maddox, et al., 2001) mapped to the Btau3.1 genome and converted to the Btau3.5x assembly using the Btau3.1 to Btau3.5x coordinate conversion chain file and the LiftOver utility. Using Perl scripts the agp file (AGP_file_specification) was built and used to generate the sequence of the virtual sheep genome assembly, the coordinate conversion chain file (for use by the LiftOver utility), and the contig and scaffold tracks for the virtual sheep genome browser (VSG).

Using the LiftOver utility and the Btau3.5x to virtual sheep genome coordinate conversion chain file, the BES and BAC-CGC mapping coordinates, and any other features mapped to the Btau3.5x bovine genome, were converted to the virtual sheep genome coordinates. Features were also transferred from the Btau3.1 genome assembly by first converting to Btau3.5x coordinates using the Btau3.1 to Btau3.5x coordinate conversion file and the LiftOver utility and then converting from Btau3.5x to vsg2.0 coordinates. Other features were mapped directly onto the virtual sheep genome using sequence alignment programs such as BLAST and BLAT with the vsg2.0 DNA sequence.

3. Results and discussion

3.1 Identification of problems with the Btau3.1 assembly of the bovine genome

The cow is the most closely related organism to sheep for which a genome assembly is available. When this project was commenced, an early draft of the bovine genome assembly Btau3.1 (Elsik, et al., 2009) was in the public domain. Since the sheep genome assembly would be built comparatively on the bovine genome, and sheep sequence contigs from the low coverage six animals at approximately 0.5 fold coverage each, were expected to be very small, the accuracy of the bovine genome assembly would determine the accuracy of the sheep assembly at all levels above that of the individual sequence contigs.

To assess the validity of this strategy the sheep BESs from the CHORI-243 library were mapped to the Btau3.1 genome assembly to identify the extent of segments of conserved synteny between the two genomes. The reader should keep in mind that the only BACs counted as being in the same organisation in the comparison genome as in the source genome (i. e. congruent) are the tail-to-tail BACs less than 200kb in length. Unexpectedly large numbers of sheep BACs, more than 17% of the BACs with both ends mapped, had both BESs positioned on the bovine Btau3.1 genome assembly within 200kb of each other, but not in the expected tail-to-tail organisation, i. e. many BACs had their two BESs mapped in the tail-to-head and head-to-head organisations (Table 1). In addition, large numbers of BACs had both BESs positioned on the same chromosome, but more than 200kb apart, the outsize groups (Table 1). The average insert size of the BACs in the sheep BAC library is 184kb (Dalrymple, et al., 2007).

Such a result would normally suggest a substantial number of intra-chromosomal rearrangements between the sheep and cattle genomes. However, almost as many, more than 14%, of bovine BACs were also not positioned as tail-to-tail BACs on the bovine Btau3.1 genome assembly (Table 1). The organisation of sheep BACs at the locations of these apparent rearrangements between the two genomes was compared with the organisation of

bovine BACs at the same locations in the genomes. Frequently clusters of tail-to-head sheep BACs overlapped with clusters of tail-to-head bovine BACs (Fig 1), suggesting that many such occurrences were in fact due to an incorrect assembly of the bovine genome, not true differences in the structures of the two genomes themselves. However, many clusters of tail-to-head sheep BACs that did not overlap with tail-to-head bovine BACs were also observed (Fig 1). These BACs probably represent rearrangements in the sheep genome relative to the bovine genome.
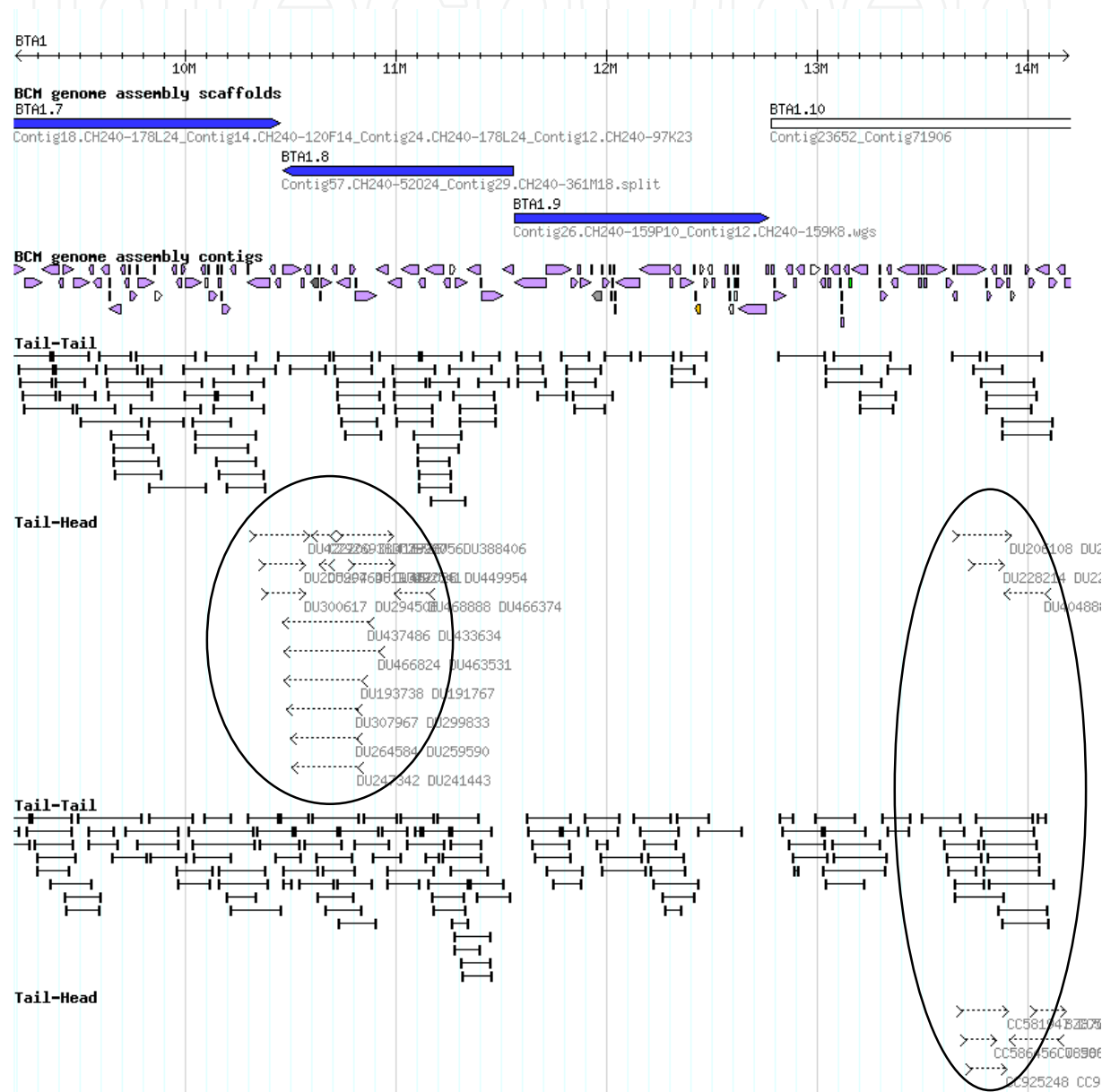

Fig. 1. Segment of chromosome one of the bovine Btau3.1 genome assembly showing the positions and orientations of sheep and cattle BACs. BCM genome assembly contigs are coloured based on the human chromosome to which they have the highest scoring match. The circles identify regions of likely inversion in the bovine and/or sheep genomes relative to the Btau3.1 genome assembly.

Genome Assembly	Btau3.1	Btau3.1	Btau3.5x	vsg2.0	vsg1.2
BAC origin	cattle	sheep	cattle	sheep	sheep
tail-to-tail (%)	86.7%	82.7%	95.63%	94.0%	89.6%
tail-to-tail outsize	2.0%	2.6%	0.4%	0.8%	1.7%
tail-to-head	3.5%	4.4%	2.6%	2.1%	2.6%
tail-to-head outsize	5.2%	7.1%	0.9%	2.1%	4.4%
head-to-head	0.4%	0.4%	0.4%	0.2%	0.3%
head-to-head-outsize	2.1%	2.8%	0.1%	0.6%	1.4%
tail-to-tail (number)	67,352	47,818	82,765	95,757	84,624
breaks		13,192		19,151	27,829
unpaired		79,172		50,142	52,663

Table 1. Mapping of cattle and sheep BACs to assemblies of the cattle and virtual sheep genomes.

3.2 Using cattle and sheep BACs to reorganise the Btau3.1 assembly of the bovine genome

The first step in the generation of the virtual sheep genome was therefore to construct the best approximation to the correct order of the bovine sequence contigs and scaffolds in the bovine genome using the bovine and sheep BACs and comparative genomics. Initially, the scaffolds in the bovine genome assembly (Btau3.1) were kept intact and scaffolds were reordered and reoriented within bovine chromosomes to minimize the number of both cattle and sheep BACs that were not in the tail-to-tail organisation. Then scaffolds apparently assigned to the wrong chromosomes on the basis of the BAC-based links to other scaffolds in the assembly were moved, including being inserted into gaps in other scaffolds guided by the mapping of the BESs. Generally these moves were also supported by the mapping of the sequence contigs to the human, dog and horse genomes (Fig. 2). In addition, scaffolds not assigned to chromosomes in Btau3.1 were included in the assembly where BACs provided unambiguous links. Finally, reordering and reorienting of contigs within the new set of ordered and reoriented scaffolds was undertaken.

Given the size of the BACs and the variation in the length of the genomic DNA contained within the BACs the correct position to insert many segments of the bovine assembly was ambiguous based solely on the BAC-end data. Throughout this process, which was mainly undertaken manually, the alignment of the bovine genome assembly contigs to the human, dog and horse genome assemblies was used in making the final decision about where exactly to insert or break scaffolds. In other words, a breakpoint between sequence contigs in an assembly scaffold was chosen that was consistent with the cattle and sheep BES data and the organisation of the human, dog and horse genomes (Fig. 2). Where conflicts between the comparative genome assemblies occurred two out of three consistent organisations were required. However, the integrity of sequence contigs was maintained throughout the process, although evidence for chimeric sequence contigs was also identified during the course of the analysis (data not shown).

To avoid ovinising the bovine genome at least one bovine BAC was required to support all reorganisations, except reordering and reorienting scaffolds within chromosomes in cases where the bovine BAC fingerprint map (Snelling, et al., 2007) also supported the
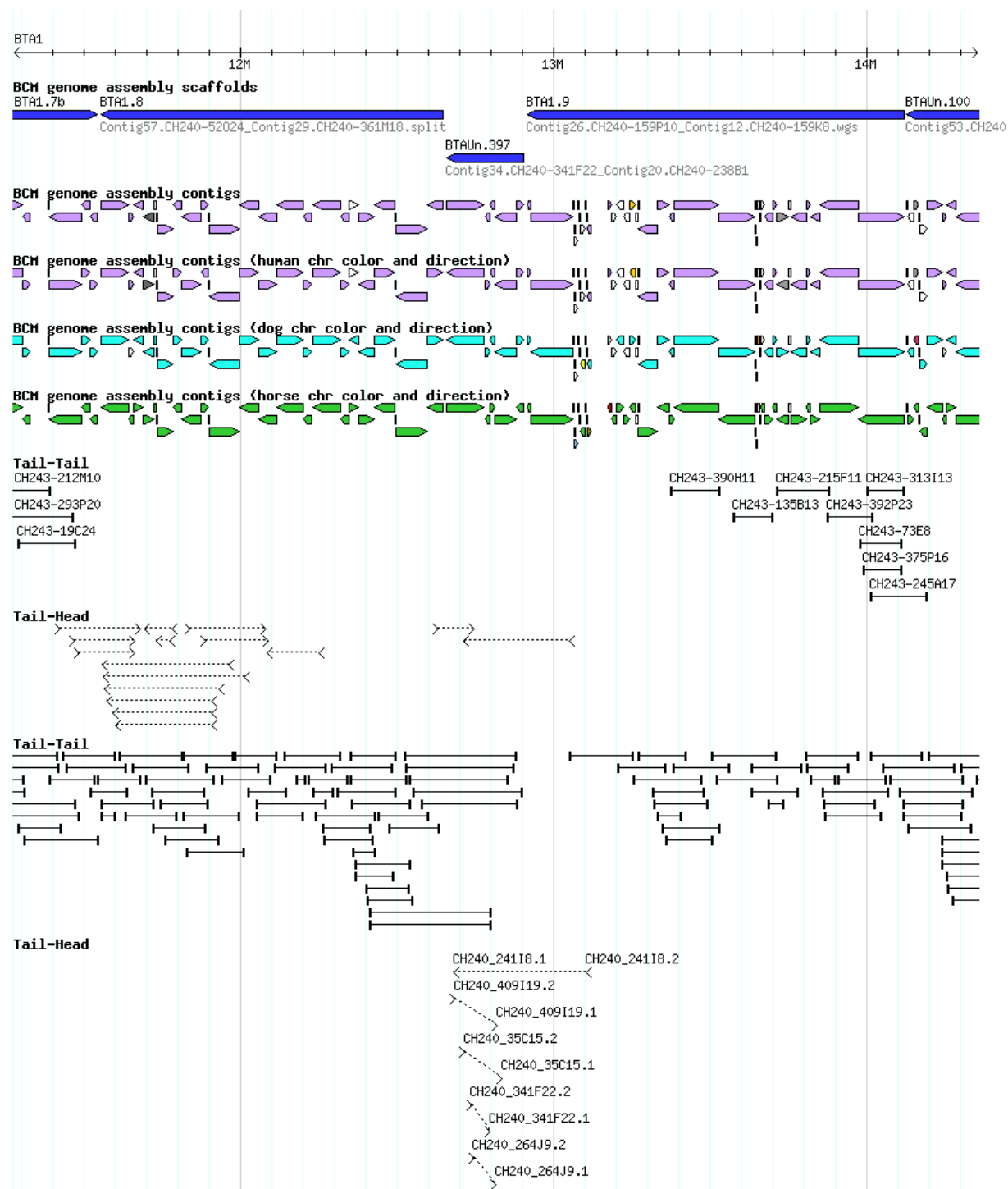
Fig. 2. Segment of chromosome one of the bovine Btau3.5x genome assembly showing the positions and orientations of sheep and cattle BACs. BCM genome assembly contigs are coloured and orientated based on the relevant species chromosome to which they have the highest scoring match The UCSC chromosome colour scheme was used (Fujita, et al., 2011).

reorganisation. This process was undertaken reiteratively to resolve any errors introduced or new links identified as the chromosome structures approached the most likely structure of the bovine genome. This revised assembly of the bovine genome based on Btau3.1 was named Btau.3.5x.

In Btau3.5x the number of bovine assembly scaffolds was reduced from 3053 scaffolds assigned to chromosomes in Btau3.1 to 537 super-scaffolds linked by the cattle and sheep BACs. Of the chromosomally assigned scaffolds in Btau3.1, 974 scaffolds were inverted, and 683 scaffolds were split into 1720 pieces, of which 710 were inverted. 14 scaffolds were moved to a different chromosome and 2192 scaffolds previously not assigned to chromosomes were incorporated into the assembly. 104 of these scaffolds were split into 233 pieces. Coverage of the genome with scaffolds assigned to chromosomes increased from 2.4 Gb to 2.77 Gb. Even after this process it is likely that there remained a number of segments of the bovine genome assembly which may not have been correctly assembled.

chromosome	scaffold	orientation	integrity
1	BTA1.1		split
1	BTAUn.418	inv	
1	BTAUn.728	inv	
1	BTA1.2		
1	BTA1.3		
1	BTA1.4		
1	BTAUn.1364	inv	
1	BTAUn.208		
1	BTA1.6		split
1	BTAUn.2125	inv	
1	BTA1.6		split
1	BTAUn.1381		
1	BTA1.5		split
1	BTAUn.1438	inv	
1	BTA1.5		split
1	BTA1.7		split
1	BTAUn.3041		
1	BTAUn.5341	inv	
1	BTA1.7		split
1	BTA1.8		

Table 2. The first twenty scaffolds of the bovine Btau3.5x assembly, scaffolds numbered BTA1.* were assigned in numerical order to chromosome 1 of the bovine genome assembly Btau3.1 build. Scaffolds numbered BTAUn.* were not assigned to a chromosome in the bovine Btau3.1 build. "inv" indicates scaffolds inverted in the Btau3.5x genome build relative to the Btau3.1 build, and "split" indicates scaffolds split in the Btau3.5x build relative to the Btau3.1 build.

3.3 Integration of the positions of sheep BESs on the Btau3.5x, dog, horse and vsg1.2 genome assemblies

We then used the virtual genome strategy (Dalrymple, et al., 2007), integrating the separate mapping of the sheep BESs to the original virtual sheep genome (v1.2), the dog and the horse genome assemblies, to maximise the positioning of sheep BESs on Btau.3.5x. There was little change in the human genome assembly over the course of the work and mapping of the sheep BACs to the human genome was captured by using the virtual sheep genome v1.2. Thus the virtual sheep genome version 2 was build on top of v1.2, rather than being a completely *de novo* version. This approach, which uses much lower specificity BLAST parameters, increased the number of sheep BACs able to be positioned on the bovine genome substantially, from 47,818 (in the initial alignments) to 95,757 in the virtual sheep genome, effectively doubling the coverage of the genome (Table 1). The number of sheep BACs able to be positioned in the tail-to-tail organisation in a genome is a complex function of the sequence coverage, assembly stage and evolutionary distance from the bovine genome. The greater distance of the dog genome appears to be partially compensated for by the more advanced state of the assembly used in this analysis. Very similar numbers of BACs were mapped in the tail-to-tail organisation to the two genomes (Table 3) with similar numbers of unique BACs (Table 4).

	bovine	horse	dog	vsg1.2
bovine	77,320			
horse	49,225	60,971		
dog	46,355	47,142	57,192	
vsg1.2	62,889	56,636	54,503	84,624

Table 3. Tail-to-tail BACs within each dataset generated by independently mapping the sheep BESs to each genome and in the intersections between each of the datasets.

genome	including vsg1.2	excluding vsg1.2
bovine	10,550	20,171
horse	701	3,035
dog	211	2,063
vsg1.2	9,204	not applicable

Table 4. Tail-to-tail BACs unique to each dataset.

The high coverage and quality of the human genome assembly and the use of the integration strategy presumably contributed to the large number of unique BACs in the tail-to-tail organisation present in vsg1.2 (Tables 3 and 4). Over and above the newer assembly of the bovine genome the inclusion of the mapping of the sheep BACs to the horse genome assembly has the biggest impact on the number of BACs assigned and on the number of BAC contigs, where fewer is better (Table 5). This is not surprising since, of the genomes used, the horse is the most closely related species to the two ruminants.

Adding the horse mapping of the sheep BESs positions to the bovine mapping of the sheep BESs positions increased the number of BACs mapped by 13,940, mainly by generating

BACs with one end directly mapped to the bovine genome and other end mapped to the bovine genome via the horse genome (Table 6). Subsequent addition of the BESs mapped via the dog genome added many fewer BACs than adding the BESs mapped via the horse genome (Table 6). The subsequent addition of the human genome data, incorporated in the vsg1.2, added slightly more BACs than the addition of the dog genome (Table 4). Thus including the dog genome had only a small impact on the improvement in the coverage of the virtual sheep genome whereas the more distant, but the better assembled/higher coverage, human genome was a useful addition to the virtual genome construction, but not unexpectedly the biggest contributions came from well assembled genomes of closely related species.

	bovine	horse	dog	vsg1.2	total
bovine	76,251				76,251
horse	13,231	709			13,940
dog	1,911	112	45		2,068
vsg1.2	3,131	284	68	55	3,538
total					95,797

Table 5. Genomes providing mapping information for the sheep BACs mapped tail-to-tail in the vsg2.0. Datasets were added in the order, bovine, horse, dog and vsg1.2.

Genome BACs positioned by	number
Both BESs vsg1.2	80,146
Both BESs bovine	13,196
One BES bovine or vsg1.2, other BES horse or dog	2431
Both BESs horse	16
One BES horse, other BES dog	5
Both BESs dog	3
	95,797

Table 6. Genomes used to position the BACs on the virtual sheep genome. Datasets were added in the order, vsg1.2, bovine, horse, and dog.

In other words, building on top of vsg1.2 and the use of a higher quality assembly of the bovine genome contributed a large number of new BACs with both ends positioned on the bovine assembly (Table 5). A large group of BACs were positioned with one end using the bovine or vsg1.2 position and the other using horse or dog. Very few BACs were positioned solely using horse and/or dog positions (Table 6). On this basis further improvement of the vsg would appear to be difficult and most likely to come from filling of gaps in the bovine genome sequence itself.

Based on the mapping of the sheep BACs to the reorganised bovine genome assembly 943 blocks of conserved synteny, defined by overlapping sheep BACs, were identified between the sheep and cattle genomes (Table 7). Assuming a genome size of 3Gb, the blocks had an average length of just over 3Mb. Although initially disappointing, even in the bovine genome assembly 537 BAC-based super-scaffolds were required to cover the complete

genome. The comparison of the number of blocks of conserved synteny identified across the different combinations of datasets demonstrates that the inclusion of additional species beyond the horse has a much greater impact on the reduction in the number of blocks of conserved synteny than it has on the total number of BACs positioned tail-to-tail. Only a 25% increase in the number of BACs, but a 56% decrease in the number blocks of conserved synteny, i. e. on average every block of conserved synteny defined based on the mapping of BACs to the bovine genome has been extended to include one adjacent block of conserved synteny.

genomes	Sheep BAC contigs
dog	2,146
horse	1,470
bovine	1,411
bovine + dog + vsg1.2	1,299
bovine + horse + dog + vsg1.2	943

Table 7. Building the virtual sheep genome.

3.4 Remaining ambiguities in the build of the bovine genome

Since there were many occasions on which there was no unambiguous basis on which to identify the correct break points in the bovine genome assembly a large number of probable inversions identified by BACs remained in the final version of the bovine genome. Most of these inversions were also supported by sheep BACs (Fig 2). In addition, whilst potentially chimeric bovine genomic sequence contigs were identified during the reassembly process, their structure has not been changed in Btau3.5x.

3.5 Construction of the virtual sheep genome (vsg2.0)

The sheep markers (sheep map version 4.7) were used to reorganise the bovine genome assembly into the vsg. In the main this involved renumbering of the bovine chromosomes, with five inverted chromosomes (or segments of chromosomes), four chromosome fusions and a single chromosome breakage (Table 8). Reordering of the segments of the bovine genome defined by the BAC comparative genome contigs (CGCs) was undertaken on four chromosomes, 7, 12, 13 and X. Apart from the X chromosome, these were local changes and involved a small number of BAC CGCs covering a small region of the genome. Given the variation in the size of BACs, and the lack of comparative data from other genomes for species specific breaks, the boundaries of such breaks could not be unambiguously identified with the data currently available. Thus no attempt was made to resolve the small potential sheep specific rearrangements within chromosomes where the break points were ambiguous and there was not sufficient marker evidence to support a change in the organisation (Fig 3).

The vsg 2.0 has been used in a number of analyses of the genome organisation of sheep and in general a high level of congruence with maps determined using other approaches has been observed (Drogemuller, et al., 2008; Wu, et al., 2008; Goldammer, et al., 2009c; Wu, et al., 2009), although the vsg 2 X chromosome build appears to contain a number of significant discrepancies (Goldammer, et al., 2009a; Goldammer, et al., 2009b).

Sheep chromosome	Cattle chromosome
OAR1	BTA3 (inv) + BTA1
OAR2	BTA8 (inv) +BTA2
OAR3	BTA11 (inv) + BTA5
OAR4	BTA4
OAR5	BTA7
OAR6	BTA6
OAR7	BTA10
OAR8	BTA9 (part)
OAR9	BTA9 (part, inv) +BTA14
OAR10	BTA12
OAR11	BTA19
OAR12	BTA16
OAR13	BTA13
OAR14	BTA18
OAR15	BTA15
OAR16	BTA20
OAR17	BTA17
OAR18	BTA21
OAR19	BTA22
OAR20	BTA23
OAR21	BTA29
OAR22	BTA26
OAR23	BTA24
OAR24	BTA25
OAR25	BTA28
OAR26	BTA27
OARX	BTAX (inv)

Table 8. High level comparison of the sheep and cattle genomes based on virtual sheep genome analysis.

3.6 Construction of the virtual sheep genome (vsg2.0) genome browser
The cattle and sheep BAC and BES locations are displayed on the chromosome overview track of the virtual sheep genome browser (VSG) allowing a quick assessment of the quality of the assembly to be made (Fig 3). In addition, the sheep virtual genome assembly was annotated with the locations of the sheep markers, SNPs on the 1536 pilot sheep SNP chip (Kijas, et al., 2009) and the Illumina Ovine SNP50 BeadChip, and human and bovine mRNA RefSeqs downloaded from the NCBI (NCBI_RefSeq).
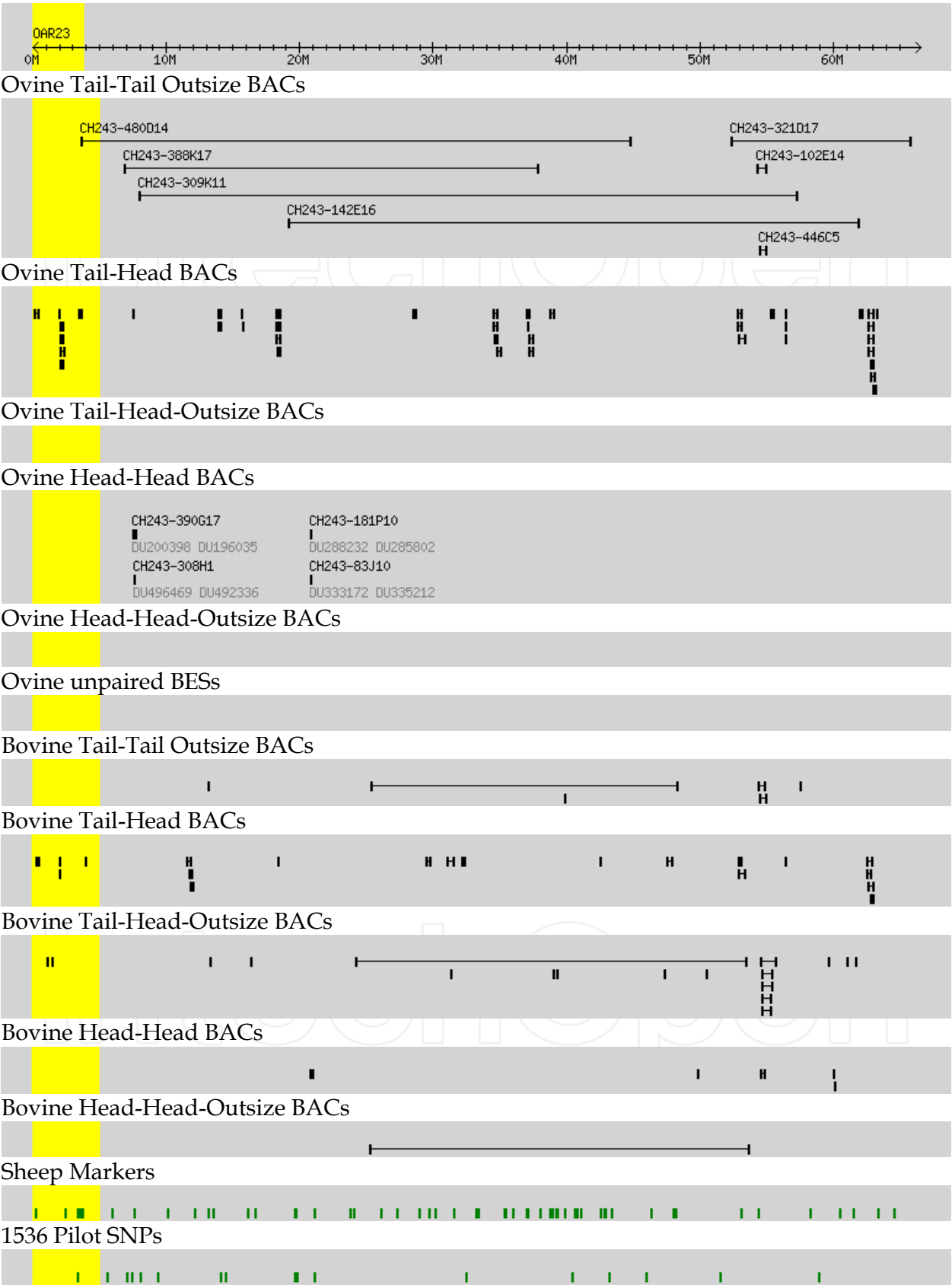
Fig. 3. Overview of chromosome OAR23 from the vsg v2 browser, displaying ovine and bovine BAC mapping, sheep linkage map markers and Pilot SNP Chip SNPs.
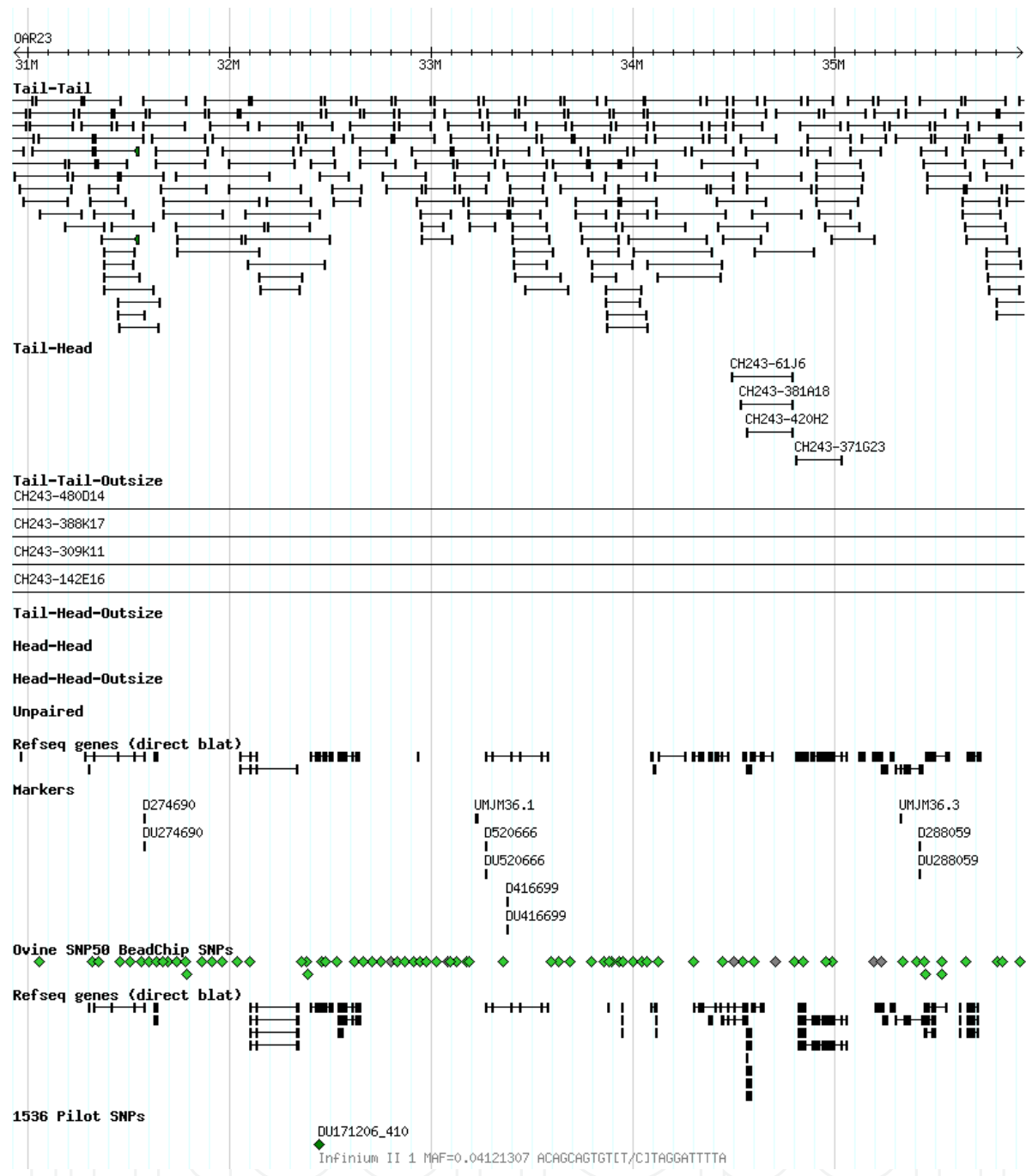
Fig. 4. A 5 Mb segment of the vsg 2 genome assembly of chromosome 23 showing the positions and orientations of sheep BACs and other tracks.

4. Conclusion

The new vsg2.0 is a significant improvement over vsg1.2, built on the human genome framework. Clearly using the genome from a closely related species and allowing the data from the species of interest to direct the process has an advantage over a very well assembled, but more distant genome. At the low resolution level down to the level of the BACs the sheep genome has a very high level of overall conserved synteny with the bovine genome structure. A number of regions of ambiguity remain, but many of these are in

regions of ambiguity of the assembly of the bovine genome and therefore await further refinement of the bovine genome assembly, or a predominantly de novo assembly of the sheep genome. However, overall it is clear that the vsg 2 makes a robust framework to assemble the large number of short contigs expected from the sequencing of the sheep genome (Archibald, et al., 2010).

Two assembled genomes from closely related species is probably the optimal balance between analysis complexity and benefit, with inclusion of a more distant, but much better assembled genome, if the genomes of closely related species are not well assembled. Thus the methods that we have described are very broadly applicable.

5. Acknowledgement

6. References

AGP File Specification (v. 1.1), Available from
     http://www.ncbi.nlm.nih.gov/projects/genome/assembly/agp/AGP_Specificatio
     n.shtml
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., Wade, C., Wang, J., Wang, W. &
     Xun, X. (2010). The sheep genome reference sequence: a work in progress. *Anim
     Genet*, Vol. 41, No. 5.pp. 449-453.
Dalrymple, B. P., Kirkness, E. F., Nefedov, M., McWilliam, S., Ratnakumar, A., Barris, W.,
     Zhao, S., Shetty, J., Maddox, J. F., O'Grady, M., Nicholas, F., Crawford, A. M.,
     Smith, T., de Jong, P. J., McEwan, J., Oddy, V. H. & Cockett, N. E. (2007). Using
     comparative genomics to reorder the human genome sequence into a virtual sheep
     genome. *Genome Biol*, Vol. 8, No. 7.pp. R152.
Drogemuller, M., Tetens, J., Dalrymple, B., Goldammer, T., Wu, C. H., Cockett, N. E., Leeb,
     T. & Drogemuller, C. (2008). A comparative radiation hybrid map of sheep
     chromosome 10. *Cytogenet Genome Res*, Vol. 121, No. 1.pp. 35-40.
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., Guigo, R., Hamernik, D. L., Kappes, S. M.,
     Lewin, H. A., Lynn, D. J., Nicholas, F. W., Reymond, A., Rijnkels, M., Skow, L. C.,
     Zdobnov, E. M., Schook, L., Womack, J., Alioto, T., Antonarakis, S. E., Astashyn, A.,
     Chapple, C. E., Chen, H. C., Chrast, J., Camara, F., Ermolaeva, O., Henrichsen, C.
     N., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Kokocinski, F., Landrum, M.,
     Maglott, D., Pruitt, K., Sapojnikov, V., Searle, S. M., Solovyev, V., Souvorov, A.,
     Ucla, C., Wyss, C., Anzola, J. M., Gerlach, D., Elhaik, E., Graur, D., Reese, J. T.,

Edgar, R. C., McEwan, J. C., Payne, G. M., Raison, J. M., Junier, T., Kriventseva, E. V., Eyras, E., Plass, M., Donthu, R., Larkin, D. M., Reecy, J., Yang, M. Q., Chen, L., Cheng, Z., Chitko-McKown, C. G., Liu, G. E., Matukumalli, L. K., Song, J., Zhu, B., Bradley, D. G., Brinkman, F. S., Lau, L. P., Whiteside, M. D., Walker, A., Wheeler, T. T., Casey, T., German, J. B., Lemay, D. G., Maqbool, N. J., Molenaar, A. J., Seo, S., Stothard, P., Baldwin, C. L., Baxter, R., Brinkmeyer-Langford, C. L., Brown, W. C., Childers, C. P., Connelley, T., Ellis, S. A., Fritz, K., Glass, E. J., Herzig, C. T., Iivanainen, A., Lahmers, K. K., Bennett, A. K., Dickens, C. M., Gilbert, J. G., Hagen, D. E., Salih, H., Aerts, J., Caetano, A. R., Dalrymple, B., Garcia, J. F., Gill, C. A., Hiendleder, S. G., Memili, E., Spurlock, D., Williams, J. L., Alexander, L., Brownstein, M. J., Guan, L., Holt, R. A., Jones, S. J., Marra, M. A., Moore, R., Moore, S. S., Roberts, A., Taniguchi, M., Waterman, R. C., Chacko, J., Chandrabose, M. M., Cree, A., Dao, M. D., Dinh, H. H., Gabisi, R. A., Hines, S., Hume, J., Jhangiani, S. N., Joshi, V., Kovar, C. L., Lewis, L. R., Liu, Y. S., Lopez, J., Morgan, M. B., Nguyen, N. B., Okwuonu, G. O., Ruiz, S. J., Santibanez, J., Wright, R. A., Buhay, C., Ding, Y., Dugan-Rocha, S., Herdandez, J., Holder, M., Sabo, A., Egan, A., Goodell, J., Wilczek-Boney, K., Fowler, G. R., Hitchens, M. E., Lozado, R. J., Moen, C., Steffen, D., Warren, J. T., Zhang, J., Chiu, R., Schein, J. E., Durbin, K. J., Havlak, P., Jiang, H., Liu, Y., Qin, X., Ren, Y., Shen, Y., Song, H., Bell, S. N., Davis, C., Johnson, A. J., Lee, S., Nazareth, L. V., Patel, B. M., Pu, L. L., Vattathil, S., Williams, R. L., Jr., Curry, S., Hamilton, C., Sodergren, E., Wheeler, D. A., Barris, W., Bennett, G. L., Eggen, A., Green, R. D., Harhay, G. P., Hobbs, M., Jann, O., Keele, J. W., Kent, M. P., Lien, S., McKay, S. D., McWilliam, S., Ratnakumar, A., Schnabel, R. D., Smith, T., Snelling, W. M., Sonstegard, T. S., Stone, R. T., Sugimoto, Y., Takasuga, A., Taylor, J. F., Van Tassell, C. P., Macneil, M. D., et al. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, Vol. 324, No. 5926.pp. 522-528.

Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T. R., Giardine, B. M., Harte, R. A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R. M., Learned, K., Li, C. H., Meyer, L. R., Pohl, A., Raney, B. J., Rosenbloom, K. R., Smith, K. E., Haussler, D. & Kent, W. J. (2011). The UCSC Genome Browser database: update 2011. *Nucleic Acids Res*, Vol. 39, No. Database issue.pp. D876-882.

Goldammer, T., Brunner, R. M., Rebl, A., Wu, C. H., Nomura, K., Hadfield, T., Gill, C., Dalrymple, B. P., Womack, J. E. & Cockett, N. E. (2009a). A high-resolution radiation hybrid map of sheep chromosome X and comparison with human and cattle. *Cytogenet Genome Res*, Vol. 125, No. 1.pp. 40-45.

Goldammer, T., Brunner, R. M., Rebl, A., Wu, C. H., Nomura, K., Hadfield, T., Maddox, J. F. & Cockett, N. E. (2009b). Cytogenetic anchoring of radiation hybrid and virtual maps of sheep chromosome X and comparison of X chromosomes in sheep, cattle, and human. *Chromosome Res*, Vol. 17, No. 4.pp. 497-506.

Goldammer, T., Di Meo, G. P., Luhken, G., Drogemuller, C., Wu, C. H., Kijas, J., Dalrymple, B. P., Nicholas, F. W., Maddox, J. F., Iannuzzi, L. & Cockett, N. E. (2009c). Molecular cytogenetics and gene mapping in sheep (Ovis aries, 2n = 54). *Cytogenet Genome Res*, Vol. 126, No. 1-2.pp. 63-76.

ISGC website, Available from http://www.sheephapmap.org

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res*, Vol. 12, No. 4.pp. 656-664.

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., Wilson, P., Ingersoll, R. G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V. H., Nicholas, F. W. & Raadsma, H. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One*, Vol. 4, No. 3.pp. e4668.

Liftover UCSC genome coordinate conversion files, Available from http://genome.ucsc.edu/cgi-bin/hgLiftOver

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., 3rd, Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A. C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J. P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, Vol. 438, No. 7069.pp. 803-819.

Liu, H., Jiang, Y., Wang, S., Ninwichian, P., Somridhivej, B., Xu, P., Abernathy, J., Kucuktas, H. & Liu, Z. (2009). Comparative analysis of catfish BAC end sequences with the zebrafish genome. *BMC Genomics*, Vol. 10, No. 592.

Maddox, J. F., Davies, K. P., Crawford, A. M., Hulme, D. J., Vaiman, D., Cribiu, E. P., Freking, B. A., Beh, K. J., Cockett, N. E., Kang, N., Riffkin, C. D., Drinkwater, R.,

Moore, S. S., Dodds, K. G., Lumsden, J. M., van Stijn, T. C., Phua, S. H., Adelson, D. L., Burkin, H. R., Broom, J. E., Buitkamp, J., Cambridge, L., Cushwa, W. T., Gerard, E., Galloway, S. M., Harrison, B., Hawken, R. J., Hiendleder, S., Henry, H. M., Medrano, J. F., Paterson, K. A., Schibler, L., Stone, R. T. & van Hest, B. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Res*, Vol. 11, No. 7.pp. 1275-1289.

NCBI RefSeq collection, Available from http://www.ncbi.nlm.nih.gov/RefSeq/index.html

Ratnakumar, A., Barris, W., McWilliam, S., Brauning, R., McEwan, J. C., Snelling, W. M. & Dalrymple, B. P. (2009). A multiway analysis for identifying high integrity bovine BACs. *BMC Genomics*, Vol. 10, No. 46.

Ratnakumar, A., Kirkness, E. F. & Dalrymple, B. P. (2010a). Quality control of the sheep bacterial artificial chromosome library, CHORI-243. *BMC Res Notes*, Vol. 3, No. 334.

Ratnakumar, A., McWilliam, S., Barris, W. & Dalrymple, B. P. (2010b). Using paired-end sequences to optimise parameters for alignment of sequence reads against related genomes. *BMC Genomics*, Vol. 11, No. 458.

Snelling, W. M., Chiu, R., Schein, J. E., Hobbs, M., Abbey, C. A., Adelson, D. L., Aerts, J., Bennett, G. L., Bosdet, I. E., Boussaha, M., Brauning, R., Caetano, A. R., Costa, M. M., Crawford, A. M., Dalrymple, B. P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C. A., Green, R. D., Holt, R., Jann, O., Jones, S. J., Kappes, S. M., Keele, J. W., de Jong, P. J., Larkin, D. M., Lewin, H. A., McEwan, J. C., McKay, S., Marra, M. A., Mathewson, C. A., Matukumalli, L. K., Moore, S. S., Murdoch, B., Nicholas, F. W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R. D., Silveri, L., Skow, L. C., Smith, T. P., Sonstegard, T. S., Taylor, J. F., Tellam, R., Van Tassell, C. P., Williams, J. L., Womack, J. E., Wye, N. H., Yang, G. & Zhao, S. (2007). A physical map of the bovine genome. *Genome Biol*, Vol. 8, No. 8.pp. R165.

Soler, L., Conte, M. A., Katagiri, T., Howe, A. E., Lee, B. Y., Amemiya, C., Stuart, A., Dossat, C., Poulain, J., Johnson, J., Di Palma, F., Lindblad-Toh, K., Baroiller, J. F., D'Cotta, H., Ozouf-Costaz, C. & Kocher, T. D. (2010). Comparative physical maps derived from BAC end sequences of tilapia (Oreochromis niloticus). *BMC Genomics*, Vol. 11, No. 636.

Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A. & Lewis, S. (2002). The generic genome browser: a building block for a model organism system database. *Genome Res*, Vol. 12, No. 10.pp. 1599-1610.

UCSC Genome Bioinformatics Website, Available from http://genome.ucsc.edu/

Virtual sheep genome browser, Available from
http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/vsheep2/

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Blocker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C., Raison, J. M., Sharpe, T., Vogel, J., Andersson, L., Antczak, D. F., Biagi, T., Binns, M. M., Chowdhary, B. P., Coleman, S. J., Della Valle, G., Fryc, S., Guerin, G., Hasegawa, T., Hill, E. W., Jurka, J., Kiialainen, A., Lindgren, G., Liu, J., Magnani, E., Mickelson, J. R., Murray, J., Nergadze, S. G., Onofrio, R., Pedroni, S., Piras, M. F., Raudsepp, T., Rocchi, M., Roed, K. H., Ryder, O. A., Searle, S., Skow, L., Swinburne, J. E., Syvanen, A. C., Tozaki, T., Valberg, S. J., Vaudin, M., White, J. R., Zody, M. C., Lander, E. S. &

Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, Vol. 326, No. 5954.pp. 865-867.

Wu, C. H., Nomura, K., Goldammer, T., Hadfield, T., Dalrymple, B. P., McWilliam, S., Maddox, J. F., Womack, J. E. & Cockett, N. E. (2008). A high-resolution comparative radiation hybrid map of ovine chromosomal regions that are homologous to human chromosome 6 (HSA6). *Anim Genet*, Vol. 39, No. 5.pp. 459-467.

Wu, C. H., Jin, W., Nomura, K., Goldammer, T., Hadfield, T., Dalrymple, B. P., McWilliam, S., Maddox, J. F. & Cockett, N. E. (2009). A radiation hybrid comparative map of ovine chromosome 1 aligned to the virtual sheep genome. *Anim Genet*, Vol. 40, No. 4.pp. 435-455.

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 Biol*, Vol. 10, No. 4.pp. R42.
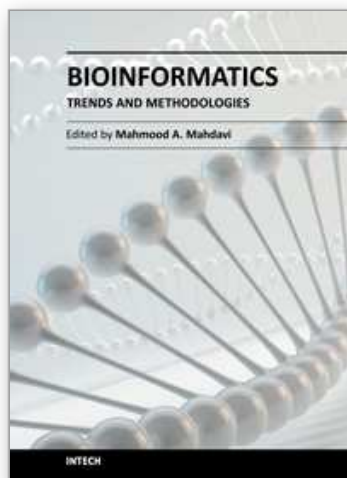
Bioinformatics - Trends and Methodologies

Edited by Dr. Mahmood A. Mahdavi

Bioinformatics - Trends and Methodologies is a collection of different views on most recent topics and basic concepts in bioinformatics. This book suits young researchers who seek basic fundamentals of bioinformatic skills such as data mining, data integration, sequence analysis and gene expression analysis as well as scientists who are interested in current research in computational biology and bioinformatics including next generation sequencing, transcriptional analysis and drug design. Because of the rapid development of new technologies in molecular biology, new bioinformatic techniques emerge accordingly to keep the pace of in silico development of life science. This book focuses partly on such new techniques and their applications in biomedical science. These techniques maybe useful in identification of some diseases and cellular disorders and narrow down the number of experiments required for medical diagnostic.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Abhirami Ratnakumar, Wesley Barris, Sean McWilliam and Brian P. Dalrymple (2011). Using Bacterial Artificial Chromosomes to Refine Genome Assemblies and to Build Virtual Genomes, Bioinformatics - Trends and Methodologies, Dr. Mahmood A. Mahdavi (Ed.), ISBN: 978-953-307-282-1, InTech, Available from: http://www.intechopen.com/books/bioinformatics-trends-and-methodologies/using-bacterial-artificial-chromosomes-to-refine-genome-assemblies-and-to-build-virtual-genomes

INTECH

open science | open minds