Journal of Heredity 2012:103(3):342–348
doi:10.1093/jhered/esr153
Advance Access publication February 7, 2012

Genome Sequence and Assembly of *Bos indicus*

Flavio C. Canavez, Douglas D. Luche, Paul Stothard, Katia R. M. Leite, Juliana M. Sousa-Canavez, Graham Plastow, João Meidanis, Maria Angélica Souza, Pedro Feijao, Steve S. Moore, and Luiz H. Camara-Lopes

From the Department of Molecular Biology, Genoa Biotecnologia SA, São Paulo, SP 04004-030, Brazil (Canavez, Luche, Leite, Sousa-Canavez, and Camara-Lopes); the Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Canada (Stothard, Plastow, and Moore); the Bioinformatics Division, Scylla Bioinformatics, Campinas, Brazil (Meidanis and Souza); and the Institute of Computing, University of Campinas, Campinas, Brazil (Meidanis, Feijao).

Address correspondence to Luiz H. Camara-Lopes, Genoa Biotecnologia SA, Rua Desembargador Eliseu Guilherme, 69/9 andar, São Paulo, SP 04004-030, Brazil, or e-mail: lhcamaralopes@genoabiotec.com.br.

Abstract

Cattle are divided into 2 groups referred to as taurine and indicine, both of which have been under strong artificial selection due to their importance for human nutrition. A side effect of this domestication includes a loss of genetic diversity within each specialized breed. Recently, the first taurine genome was sequenced and assembled, allowing for a better understanding of this ruminant species. However, genetic information from indicine breeds has been limited. Here, we present the first genome sequence of an indicine breed (Nellore) generated with 52X coverage by SOLiD sequencing platform. As expected, both genomes share high similarity at the nucleotide level for all autosomes and the X chromosome. Regarding the Y chromosome, the homology was considerably lower, most likely due to uncompleted assembly of the taurine Y chromosome. We were also able to cover 97% of the annotated taurine protein-coding genes.

Key words: Bos indicus, *bovine, genome, Nellore, sequencing*

Domestic cattle comprise more than 800 breeds that are artificially grouped into 2 species, *Bos taurus* (taurine) and *B. indicus* (indicine or zebu). Several phenotypic differences have been described between indicine and taurine cattle (Bradley and Magee 2006), the most notable of which is a prominent hump over the shoulder and neck in indicine cattle. Important physiological advantages of indicine cattle over taurine cattle include greater resistance to heat, less susceptibility to ticks and gastrointestinal parasites, and a lower metabolic rate and nutrient requirements.

Origins of indicine and taurine are controversial among the archaeologists who have reached no consensus about their centers of domestication. However, analysis of mitochondrial DNA sequences has shown divergence of >250 thousand years between these 2 types, indicating at least 2 distinct centers (Bradley et al. 1996). Y-chromosome data also support more than one center and point to genetic introgression as an important event for modern cattle evolution (Pérez-Pardal et al. 2010).

Among the major meat and milk producing countries, the global distribution of herds shows the predominant use of taurine in European countries and Argentina, whereas indicine cattle predominantly form Indian, Brazilian, southern United States, northern Australia, and southern China herds.

Cattle breeds selected for beef and dairy production constitute major nutritional sources for human populations. As ruminants, they are specialized in converting low-grade forage into energy-rich fat, milk, and muscle. This importance has led humans to exert strong selection pressures on these breeds by forming herds using a small number of founder animals, which exhibit the phenotype of interest. Side effects of this process are significant reductions in genetic diversity (Taberlet et al. 2009) leading to metabolic, reproduction, and health problems (Rauw et al. 1998). Moreover, less productive breeds are usually neglected, which in turn can lead to their extinction, as it has already for 209 cattle breeds (Food and Agriculture Organization of the United Nations 2007). These developments have profound impacts on the species, especially considering that wild cattle no longer exist and that all surviving genetic diversity is now present in domestic animals. Therefore, a better understanding of cattle genetics could help us reduce some of these undesirable effects.

Recently, DNA taken from a Hereford taurine breed sire and daughter (Dominette 01449) was sequenced (Elsik et al. 2009), assembled by 2 different groups (Burt 2009), and annotated (Reese et al. 2010; Childers et al. 2011). Although based on the same sequence data generated by The Bovine Genome Sequencing and Analysis Consortium, 2 different genome assemblies were generated using different approaches. BCM4 (Liu et al. 2009) was assembled by combining whole-genome shotgun reads, and bacterial artificial chromosomes sequences leading to a 2.54 Giga base-pair (Gbp) assembled genome. UMD2 (Zimin et al. 2009) utilized paired-end sequence information, mapping data, and the human genome to assemble contigs, leading to a 2.62 Gbp sequence that includes a tentative chromosome Y. These were important early steps toward revealing the genetic characteristics of ruminant species (Moore 2009).

Here, we present the first genome of the indicine species and compare it with both previously published assemblies of a taurine genome. This may help to provide insights into the underlying genes and variants contributing to these traits. Furthermore, we compared the protein-coding genes annotated in the *B. taurus* genome with those present in our assembled genome. The animal used was a male Brazilian Nellore, a breed imported from India mainly in the first half of the 20th century. In Brazil, Nellore was developed for beef production and represents the most commonly used breed due to its tolerance to high temperatures.

Materials and Methods

Animal Selection and DNA Extraction

We used a Nellore sire born on 14 December 2004 and registered at "Associação Brasileira de Criadores de Zebuino (ABCZ)" as QUIL7308 (Figure 1). The selected animal is rated in the top 0.1% for some phenotypic traits of economic impact. Currently, the animal is kept at "CRV Lagoa" (http://www.crvlagoa.com.br), where his semen is collected and marketed throughout the country.

Figure 1. The Nelore bull QUIL7308 that was used as DNA source for the first indicine genome sequence.

The peripheral blood mononuclear cells (PBMC) samples were collected from the jugular vein using EDTA as anticoagulant. Then, genomic DNA was extracted from PBMCs using the QIAamp DNA Mini Kit (QIAGEN) according to manufacturer's instructions. Lastly, genomic DNA was recovered in Tris–HCl pH 8.0 at 50 mM and quantified using Nanodrop. Approximately, 500 μg of genomic DNA was used in the construction of libraries and subsequent sequencing.

Genome Sequencing and Assembly

Approximately 4 billion usable reads were generated on an ABI SOLiD sequencing platform. Of these, 2.2 billion have read lengths of 50 nt and 1.8 billion have read lengths of 25 nt. The reads were organized in 6 libraries: 2 single-read 50 nt-read libraries (with 0.7 billion reads altogether), 2 mate-paired 25 nt-read libraries (with a total of 1.8 billion reads), and 2 mate-paired 50 nt-read libraries (totaling 1.5 billion reads). The insert sizes used in sequencing were: 110 bp for single-read libraries, 1.3–2.5 Kb for the 25 nt mate-paired libraries, and 1–2 Kb for the 50 nt mate-paired libraries. Table 1 summarizes read distribution in the libraries. Quality control removed 2% of the reads, which for at least one position had no base set. The remaining reads provided 52X coverage of the indicine genome, with 58.07% of these reads having median quality between 20 and 35 (never exceeding 35) and 17.69% having median quality less than 10.

Read alignment was executed against the 2 *B. taurus* reference genomes in parallel, both deposited at NCBI: Project ID:13366 (Liu et al. 2009), BCM4 (Baylor College of Medicine Human Genome Sequencing Center) with 30 chromosomes (29 + X chromosome); Project ID:33843 (Zimin et al. 2009), UMD2 (University of Maryland assembly of *B. taurus*) with 31 chromosomes (29 + X + Y chromosome). Considering that the sequenced genome in this work is that of a male, an additional Y chromosome reference was used for Y chromosome assembly: Project ID:20275 (Skaletsky, H, Dugan-Rocha, S, Larracuente, AM, Ding,Y, Brown, LG, Khan, Z, Pyntikova, T, Villasana, D, Koutseva, N, Buhay, C, Cho, G, Jiang, H, Wilczek-Boney, KB, Worley, KC, Hughes, J, Muzny, DM, Page, D and Gibbs, RA, Baylor College of Medicine, unpublished 2011), Chromosome Y sequencing.

Alignment was performed with BWA (Burrows–Wheeler Alignment tool) (Li and Durbin 2009), in 3 steps: 1) reference genome indexing; 2) find read coordinates in the reference; and 3) generate the alignments in Sequence Alignment Map (SAM) format. Steps (1) and (2) included the -c option, suitable for SOLiD reads. In step (2), after several tests, we included options -c -t -m 7 and -R 25 (this last one for 50 bp libraries only). The -t option is for multithreading. In step (3), -t is not possible, but we optimized memory use with the -P option. The package SAMtools (Li et al. 2009) was used to process the SAM files in order to visualize the alignments and obtain the consensus sequences for each library. To build a consensus including all libraries, the SAM files were concatenated and converted to a BAM file, which was fed to the pileup SAMtool.

Table 1 Number of reads and insert sizes for all 6 libraries constructed and used for the indicine genome sequencing

Library	Number of reads (million)	Insert size
Single1	340	110 bp
Single2	360	110 bp
Paired_25_1	928	1.3–1.5 Kb
Paired_25_2	846	1.5–2.5 Kb
Paired_50_1	776	1–2 Kb
Paired_50_2	774	1–2 Kb

To further evaluate the assembly, we used base qualities of the generated contigs, and also the N50 metric. Base qualities, as reported by BWA, are based on the PHRED quality scale (Ewing and Green 1998), a de facto standard for genome assemblies. The N50 metric is the size of the smallest contig that has to be considered to cover at least 50% of a sequence (in our case, each chromosome). High values of N50 indicate better assemblies.

Protein-Coding Gene Comparison

The *B. indicus* assembled genome was used to search for coding genes (proteins) annotated in the *B. taurus* BCM4 assembled genome (Reese et al. 2010; Childers et al. 2011). The 29 autosomal chromosomes and the X chromosome were compared with a total of 20 151 NCBI *B. taurus* coding genes, extracted from the *B. taurus* annotation. Each *B. taurus* coding gene was initially compared using tblastn (Altschul et al. 1997) with the corresponding *B. indicus* chromosome. If no significant hits were found, a new tblastn comparison was performed against the entire *B. indicus* genome. All tblastn runs used the default parameters, except that we disabled masking of low complexity regions. A challenge with the tblastn approach is that a *B. taurus* coding gene may still produce a very good hit even if its sequence is missing from *B. indicus*, due to the presence of paralogs or conserved domains. To complement the analysis, we also looked for coding genes in *B. taurus* contained in regions not covered by the *B. indicus* reads.

Codon usage of *B. indicus* was comparable to that of *B. taurus* (BCM4 reference). For BCM4, DNA-coding sequences were obtained from NCBI, and the frequency of each of the 64 codons was observed. For *B. indicus*, codon usage was computed based on regions used in our annotation.

Results and Discussion

The *B. indicus* genome alignment against reference BCM4 (Liu et al. 2009) resulted in 319 222 contigs for the 30 chromosomes (29 autosomes and X chromosome), covering 93% of the reference. If restricted to contigs larger that 200 bp, 218 370 contigs were obtained for the 30 chromosomes (29 autosomes and X chromosome), covering 92.6% of the reference. The *B. indicus* genome alignment against reference UMD2 (Zimin et al. 2009) resulted in 372 804 contigs for the 30 chromosomes (29 autosomes and X chromosome), covering 99.1% of the reference. If restricted to contigs larger that 200 bp, the outcome was 265 901 contigs placed onto the 30 chromosomes, covering 98.8% of the reference. As expected, requiring contigs larger than 200 bp yielded a slightly lower, although probably more accurate, coverage. Tables 2 and 3 summarizes and compares the coverage against both reference assemblies for contigs larger than 200 bp.

The average quality of assembled contigs, considering contigs larger than 200 bp only, reaches around 100 points in PHRED scale, meaning less than one error per 10 billion bases, for the 29 autosomes of both assemblies. Chromosome X yielded lower average base qualities (72–75), and for chromosome Y, we had even poorer results (less than 20), which was partly expected, due to the unfinished status of the Y.

Regarding the N50 metric, related to the ability of constructing long stretches from the reads, we obtained results averaging about 26 000 for autosomes in the BCM4 assembly, and slightly lower near 23 000 for autosomes in the UMD2 assembly. For chromosome X the values were about half of these, and for Y, 2 orders of magnitude lower. The autosomal values are surprisingly close, probably because these are reference assemblies. The values per se seem to be lower than in similar projects, which may mean that we were too strict in our assembly thresholds.

Supplementary Table 1 presents average size and coverage for all contigs.

Figure 2 illustrates the total contig sizes for each chromosome against BCM4 reference (red) and UMD2 reference (yellow).

Overall, the results using the 2 assemblies are similar and together suggest a high level of similarity between the taurine and the indicine genomes, although differences are definitely apparent. Additionally, sequences possibly present in *B. indicus* but absent in *B. taurus* will not be revealed by this approach. Assembly-specific differences are easily noted almost uniformly for each and every autosome (Figure 2), which is not surprising, given that the assemblies were produced using different approaches. Various metrics suggest that the UMD2 assembly is more accurate (Zimin et al. 2009) than the BCM4 assembly, a fact that could explain the better coverage obtained in this work using UMD2.

Worth noting is also the discrepancy in size for the assembly of the X chromosome between both reference genomes. UMD2 assigned 136 Mbp of sequence to the X chromosome, whereas 83 Mbp was assigned by BCM4. Interestingly, the trend in coverage rate observed for the 29 indicine autosomes remained in the same range for chromosome X (Tables 2 and 3), regardless of one being 56% of the other's size.

One and a half billion indicine reads were not able to align against the BCM4 taurine 30 chromosomes (29 autosomes plus X). These were used for the assembly of the Y chromosome. Alignment against the reference Chromosome Y provided by Baylor College of Medicine (unpublished) resulted in 257 919 contigs and 65% coverage, whereas alignment against the Y reference in

Table 2 Quality indicators on reference assemblies performed against BCM4

	BCM4				
Chromosome	Number of contigs	Average contig size	Percentage of coverage	N50	Average contig quality (PHRED scale)
1	17 890	8376.84	92.84	27 534	102
2	15 391	8627.94	94.21	28 739	102
3	14 241	8378.54	93.12	29 226	104
4	14 570	8083.93	94.46	27 243	102
5	15 690	7489.9	93.29	25 992	101
6	14 579	7759.08	92.02	26 645	102
7	13 815	7624.07	93.86	26 120	102
8	12 563	8625.18	92.48	28 088	103
9	12 294	8250.97	93.66	27 382	101
10	11 946	8415.24	94.37	28 917	102
11	14 262	7277.87	93.87	27 988	102
12	11 431	7064.98	94.28	25 067	100
13	8104	9838.33	94.27	30 071	104
14	8997	8582.87	94.68	27 702	102
15	10 814	7319.64	93.12	25 681	101
16	8809	8215.71	92.70	26 974	103
17	9347	7600.16	92.64	24 872	101
18	8887	6877.07	92.40	24 211	102
19	8092	7528.07	93.03	24 719	101
20	9200	7697.69	93.14	26 254	101
21	8907	7307.11	93.65	26 360	101
22	6492	9034.76	94.60	29 376	104
23	7388	6786.52	93.71	24 794	101
24	7812	7837.13	93.95	26 464	101
25	6307	6552.92	93.57	22 898	101
26	6496	7440.77	92.98	27 036	101
27	6944	6522.05	92.60	25 466	100
28	5042	8553.82	93.36	26 651	103
29	7455	6430.35	91.68	22 591	100
X	15 464	5334.12	92.88	14 427	75
All	319 229	7726.66	92.60	26 236	101
Y	35 042	430.13	37.3	201	14

Only contigs larger than 200 bp are considered. For each chromosome, we list its number, total size, average size of mapped contigs, coverage by mapped contigs, N50 measure (smallest contig size threshold needed to cover at least 50% of the chromosome), average base quality of mapped contigs (PHRED scale, as reported by BWA). Chromosome Y appears after the totals because it was mapped separately from the other chromosomes.

UMD2 resulted in 81 734 contigs and 16.1% coverage. For contigs larger than 200 bp, the results were 35 042 contigs with 37% coverage (contig size of 15 072 505 bp) and 2 594 contigs with 2.6% coverage (contig size of 1 051 843 bp), respectively. The effort of completing the bovine chromosome Y is an ongoing and still unfinished project, and the differing coverage and in general low coverage observed in this work likely reflect the incomplete nature of the reference assemblies. With that in mind, a proper comparison and more conclusive evaluation of the Y chromosome information awaits to be completed.

Almost all of the *B. taurus* protein-coding genes (99.98%) were found in the *B. indicus* genome with the default tblastn parameters. The sequence comparisons conducted revealed just 2 proteins with no hits whatsoever: gi|297466360|ref|-XP_002704465.1| (PREDICTED: hypothetical protein) on chromosome 5 and gi|118151338|ref|NP_001071601.1| (beta-defensin) on chromosome 27. More definitive experiments will be required to establish that the coding sequences for these proteins are truly missing in *B. indicus*. The coverage (% of length) for each reference protein–coding gene was calculated, as well as the arithmetic mean coverage and the weighted mean coverage (weighted by the length of the proteins) for each chromosome, which can be viewed as the exon percent coverage across the entire chromosome (Table 4). These values show that the protein-coding genes known in the cattle genome are well represented in our assembly of the *B. indicus* genome, with 98.24% of exon coverage overall. In addition, looking at the coverage of *B. taurus* protein-coding genes by the *B. indicus* assembly, we found out that 97.03% of the genes have at least 90% of their length covered by our reads. This number rises to 99.43% if considered genes covered at least 50% by *B. indicus* reads.

With respect to codon usage, comparison between *B. indicus* and *B. taurus* BCM4 showed that, for most amino acids, the difference between relative preferences for a particular codon is at most 5%. The exceptions are:

Table 3 Quality indicators on both reference assemblies performed against UMD2

	UMD2				
Chromosome	Number of contigs	Average contig size	Percentage of coverage	N50	Average contig quality (PHRED scale)
1	13 290	11 806.16	99.09	25 345	102
2	11 199	12 135.75	99.16	25 912	102
3	10 334	11 640.96	99.07	26 514	103
4	10 843	11 047.17	99.13	24 869	102
5	11 532	10 395.16	98.92	23 994	102
6	11 546	10 239.99	98.97	24 608	101
7	10 359	10 666.55	98.10	23 537	102
8	9605	11 696.97	99.09	25 750	103
9	9031	11 600.06	99.10	25 666	102
10	9150	11 293.48	99.07	26 136	102
11	9545	11 133.18	99.03	24 421	102
12	12 158	7365.83	98.23	20 105	98
13	7153	11 675.99	99.14	25 227	103
14	8350	10 035.14	98.99	24 492	102
15	7702	10 943.9	98.82	24 093	102
16	7887	10 245.42	98.88	22 714	102
17	7374	10 089.38	98.99	22 649	102
18	7312	8911.6	98.72	21 975	101
19	6569	9647.76	98.94	21 804	100
20	6724	10 614.36	99.07	23 391	101
21	6979	10 145.82	98.89	23 455	101
22	5331	11 432.92	99.21	25 745	103
23	5439	9541.05	98.79	21 428	101
24	5908	10 519.21	99.10	22 702	101
25	5108	8313.79	98.98	20 154	100
26	5043	10 144.42	98.99	23268	102
27	4988	8979.67	98.64	21 247	102
28	4261	10 771.34	99.10	22 952	102
29	6256	8126.86	98.71	19 470	101
X	28 925	5044.31	98.04	11 160	72
All	265 901	9895.92	98.80	22 972	100
Y	2594	405.49	2.6	200	17

Only contigs larger than 200 bp are considered. For each chromosome we list its number, total size, average size of mapped contigs, coverage by mapped contigs, N50 measure (smallest contig size threshold needed to cover at least 50% of the chromosome), average base quality of mapped contigs (PHRED scale, as reported by BWA). Chromosome Y appears after the totals because it was mapped separately from the other chromosomes.

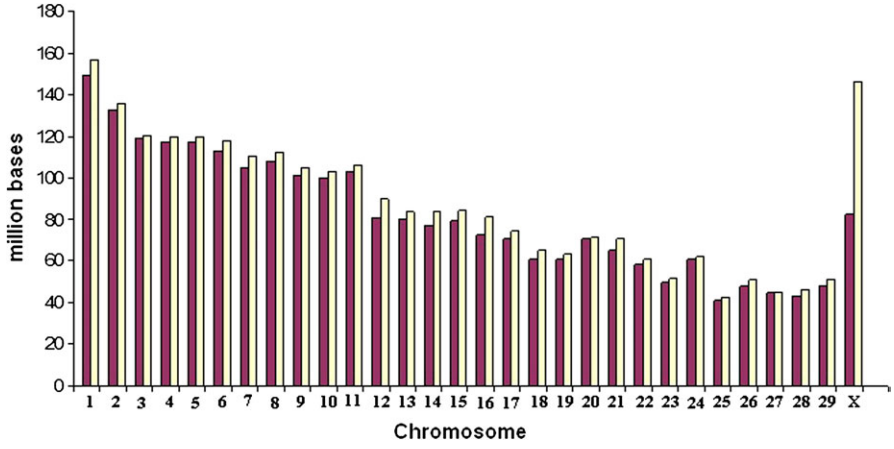

Figure 2. Mapping of all contigs (larger than 200 bp) from the indicine bull QUIL7308 against autosomal and X chromosomes of taurine genomes BCM4 and UMD2. Bars show the total size of indicine contigs mapped against BCM4 and UMD2, in red and yellow, respectively. Overall, larger contigs were obtained in the mapped against UMD2, although values and patterns are very similar in both reference assemblies. In addition, X chromosome comparisons generate different results due to size divergences between UMD2 and BCM4.

Table 4 Statistics of protein coverage

Chromosome	*B. taurus* proteins	Proteins found	WM (%)	AM (%)	MinC (%)	MaxC (%)	MedC (%)
1	814	814	98.49	98.37	35.42	100	100
2	854	854	98.32	97.98	17.46	100	100
3	1182	1182	98.29	98.08	39.41	100	100
4	698	698	97.79	97.84	13.07	100	100
5	1132	1131	98.18	98.32	58.82	100	100
6	579	579	97.52	96.92	12.41	100	100
7	1175	1175	98.4	98.22	35.71	100	100
8	657	656	98.46	98.14	28.74	100	100
9	464	464	98.39	98.11	60.27	100	100
10	936	936	98.22	98.09	47.14	100	100
11	916	916	98	98.08	29.6	100	100
12	360	360	98.5	97.91	60.4	100	100
13	700	700	97.92	97.66	32.17	100	100
14	401	401	98.26	97.85	58.33	100	100
15	923	923	98.61	98.73	39.94	100	100
16	525	525	98.57	98.25	40.45	100	100
17	547	547	97.84	97.85	10.04	100	99.98
18	1112	1112	98.13	98.2	41.18	100	100
19	1163	1163	98.64	98.46	11.71	100	100
20	316	316	98.53	98.32	22.47	100	100
21	455	455	98.33	97.89	34.43	100	100
22	536	536	98.52	98.33	42.38	100	100
23	670	669	97.94	97.84	19.34	100	100
24	291	291	98.32	97.95	64.35	100	100
25	685	685	97.7	97.49	32.4	100	100
26	369	369	98.71	98.55	42.86	100	99.89
27	248	247	97.65	97.12	20.86	100	99.68
28	264	264	98.44	97.93	63.67	100	100
29	584	584	98.2	97.87	34.57	100	100
X	595	595	98.16	98.14	18.02	100	100
All	20151	20147	98.24	98.07	10.04	100	100

For each chromosome, we have: chromosome number; number of *Bos taurus* proteins in the official BCM4 annotation; number of *Bos indicus* proteins found in our study; weighted mean of coverage (WM%), which is percent coverage for each protein weighted by the protein's length; arithmetic mean of coverage (AM%); minimum coverage (MinC%); maximum coverage (MaxC%); and median coverage (MedC%).

B. taurus prefers the GTG Valine codon 8.5% above *B. indicus* and the TGA Stop codon 14% above *B. indicus*.

Conclusions

Two interesting conclusions can be drawn. First, the approach of assembling an entire genome using very short DNA reads (25 and 50 nt) from small DNA insertion libraries (110 bp to 2 kb) can produce satisfactory results, as indicated here by the high coverage obtained. However, this approach is only feasible if the genome of a very closely related species is already available as a template. Second, although these 2 breeds were probably domesticated independently and following different migration routes (Caramelli 2006), they have been separated from each other by a short period of time (200 KYA–1 Ma). Hence, if confirmed by further experiments, the 2 taurine proteins absent in the indicine genome would be very intriguing.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

Genoa Biotecnologia SA (GBSA1308), Sao Paulo, Brazil.

Acknowledgments

Special thanks to Dr Vilson A. Simon, General Manager, MSD Saude Animal, Brazil for support and encouragement and to Mrs. Alice Ferreira for kindly providing the biological source used for sequencing, the Nellore sire QUId. This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AGFL00000000. The version described in this paper is the first version, AGFL01000000. The authors have no conflicts of interest to declare.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 5:3389–3402.

Bradley DG, Machugh DE, Cunningham P, Loftus RT. 1996. Mitochondrial diversity and the origins of African and European cattle. Proc Natl Acad Sci U S A. 93:5131–5135.

Bradley DG, Magee DA. 2006. Genetics and the origins of domestic cattle. In: Zeder MA, Bradley DG, Emshwiller E, Smith DB, editors.

Documenting domestication: new genetic and archaeological paradigms. 1st ed. Berkeley (CA): University of California Press. p. 317–318.

Burt DW. 2009. The cattle genome reveals its secrets. J Biol. 8:36.

Caramelli D. 2006. The origins of domesticated cattle. J Hum Evol. 21:107–122.

Childers CP, Reese JT, Sundaram JP, Vile DC, Dickens CM, Childs KL, Salih H, Bennett AK, Hagen DE, Adelson DL, et al. 2011. Bovine Genome Database: integrated tools for genome annotation and discovery. Nucleic Acids Res. 39:D830–834.

Elsik CG, Tellam RL, Worley KC. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 324:522–528.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. Genome Res. 8:186–194.

Food and Agriculture Organization of the United Nations (FAO). 2007. Report by the Commission on Genetic Resources for Food and Agriculture. Rischkowsky B, Pilling D, editors. The state of the world's animal genetic resources for food and agriculture. Rome (Italy) Chapter Methods for Conservation, p. 467–468.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 25:2078–2079.

Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, et al. 2009. Bos taurus genome assembly. BMC Genomics. 10:180.

Moore SS. 2009. The bovine genome sequence—will it live up to the promise? J Anim Breed Genet. 126:257–258.

Pérez-Pardal L, Royo LJ, Beja-Pereira A, Chen S, Cantet RJC, Traoré A, Curik I, Sölkner J, Bozzi R, Fernández I, et al. 2010. Multiple paternal origins of domestic cattle revealed by Y-specific interspersed multilocus microsatellites. Heredity. 105:511–519.

Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livest Prod Sci. 56:15–33.

Reese JT, Childers CP, Sundaram JP, Dickens CM, Childs KL, Vile DC, Elsik CG. 2010. Bovine Genome Database: supporting community annotation and analysis of the Bos taurus genome. BMC Genomics. 11:645.

Taberlet P, Coissac E, Pansu J, Pompanon F. 2009. Conservation genetics of cattle, sheep, and goats. C R Biol. 334:247–254.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, et al. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10(4):R42.

Received September 9, 2011; Revised December 21, 2011; Accepted December 21, 2011

Corresponding Editor: James E. Womack