

Polymorphism and mobilization of transposons in *Bos taurus*

Bernt Guldbbrandtsen *

*bernt.guldbbrandtsen@agrsci.dk, Quantitative Genetics and Genomics, Dept. of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark

ABSTRACT: The bovine genome assembly was explored to detect putative retrotransposon sequences. In total 87,310 such sites were detected. Four breeds of dairy cattle were examined with respect to the presence, segregation or complete absence of the putative retrotransposon. A total of 10,983 sites were polymorphic. Some were absent in one or more of the breeds, including 680 which were absent in the Jersey breed, but only 13 were completely absent in all dairy breeds. The fact that so many of the putative retrotransposons are polymorphic, suggests that retrotransposition has been happening at least until the very recent past.

Keywords:

Bos taurus

Retrotransposon

Variation

Introduction

The existence of genetic elements able to move and copy themselves about in the genome, mobile genetic elements has been recognized for a long time (McClintock 1956). Transposons and transposon-like elements constitute an estimated 44% of the human genome although only few of these remain functional (Mills, Bennett, Iskow et al. 2007). Two families of mobile genetic elements, Alu1 and LINE1 constitute 10.0% and 14.6% of the total human genome, respectively (Smit 1996).

Retrotransposons are known to play roles in genome evolution (e.g. Janoušek, Karn, and Laukaitis 2013). Kazazian Jr (1998) gives an estimate that 10% of spontaneous mutations are caused by retrotransposition events. Recently, Kadri, Sahana, Charlier et al. (2014)

showed that long interspersed nuclear elements are involved in the mutagenesis of a deletion acting as a major recessive QTL affecting fertility in Nordic Red cattle.

Almeida, Silva, Silva Jr. et al. (2007) and Adelson, Raison, and Edgar (2009) examined gene sequences of *Bos taurus* and found that transposable genetic elements make a large contribution to bovine gene structure. Particularly they found that insertions into intron sequences were abundant.

The easy availability of whole genome sequencing data makes it possible to study the variability of the occurrence of LTR retrotransposons – both within and between populations. Ellinghaus, Kurtz, and Willhoeft (2008) introduced a sensitive tool to detect putative members of a class of retrotransposons, LTR flanked retrotransposons with good specificity (100% in yeast, 72% in drosophila).

Here evidence is presented confirming the widespread occurrence in the genome of *Bos taurus* of putative retrotransposons in the form of LTR flanked regions. Also, evidence for that a substantial proportion of these putative retrotransposons are polymorphic with respect to their presence or absence in the genomes of animals belonging to modern dairy cattle breeds. This suggests extensive retrotransposition activity at least until the recent past.

Materials and Methods

Whole-genome sequencing data at a depth of 10X or better were available for 172 animals. All animals belonged to dairy breeds. There were 46 Holstein (HOL), 27 Jersey (JER), 83 Red Dairy Cattle (RDC) and 16 Brown Swiss (BSW). Sequences were processed according to

found at <http://www.1000bullgenomes.com/>.

To detect putative retrotransposons `Ltrharvest` (part of `GenomeTools`) by Ellinghaus, Kurtz, and Willhoeft (2008) was run with default parameters against the UMD 3.1 assembly of the *Bos taurus* genome (Zimin, Delcher, Florea et al. 2009).

For each `ltrharvest` hit each BAM file was examined for read pairs that spanned the LTR in question. For this to happen one of the reads in a pair must map to the left and the other one to the right of the `ltrharvest` hit. The BAM files are scanned for reads where one read maps to the left of the position of the `ltrharvest` hit and where the length of the mapped region spanned by the reads is sufficient to hold the `ltrharvest` hit. These reads were counted. The following filtering was applied: 1) If any animal had more than 100 read pairs satisfying the criteria the hit was discarded. 2) If any animal had more than 30 reads pairs satisfying the criteria the breed was discarded. 3) Any `ltrharvest` hit with retained in less than five animals in a breed was discarded for this breed (“NA”). 4) Any animal with zero or one read satisfying the criteria was classified as carrying the sequence of the `ltrharvest` hit. 5) Any animals with more than one read satisfying the criteria was classified as not carrying the read on at least one of its chromosome copies. A breed was classified as “polymorphic” is both carrying and non-carrying animals were present, “not present” is only non-carrying, “retained” if no non-carriers were found and “NA” if five or less animals passed filtration.

An `ltrharvest` hit was considered polymorphic if:

- No breed classified was “NA”, and
- At least one breed was classified as either polymorphic or “not present”.

Results

After filtering this produced a total 87,310 hits. The distribution of hits across the chromosomes is shown in table 1. Typical densities were in the range between 30 and 60 hits per Mb with extremes at 1 per Mb and 167 per Mb. Among the hits 932 have quality scores greater than 99, 6,800 have scores greater than 95 and only 330 have scores less than 80.

Table 1: Distribution of putative LTR retrotransposons across the *Bos taurus* genome. Numbers are counts of hits (Cnt) for each *Bos taurus* chromosome (Chr).

Chr	Cnt	Chr	Cnt	Chr	Cnt
1	5,810	11	3,189	21	2,204
2	4,647	12	3,142	22	1,627
3	4,068	13	2,294	23	1,500
4	4,142	14	2,850	24	1,852
5	3,953	15	2,800	25	1,040
6	4,607	16	2,540	26	1,624
7	3,623	17	2,301	27	1,513
8	3,988	18	1,719	28	1,421
9	3,720	19	1,528	29	1,592
10	3,384	20	2,439	X	6,193

Out of the 87,310 `ltrharvest` hits, 70,046 were present in all animals in all the dairy breeds examined. Of the sites detected in the Hereford genome, 736 are absent in at least one of the dairy breeds, 13 were absent from all dairy animals in all the breeds, while 2,514 sites all four breeds were polymorphic in at least one dairy breed. The numbers of putative LTR retrotransposons were classified with respect to presence and polymorphism are shown in table 2.

The relatively high number of “NA” results for BSW probably reflects that the number of animals sequenced for this breed was the smallest.

Table 2: Number of putative LTR transposons that were present in all individuals in a breed (Retained), missing in at least one chromosome in all animals in a breed (Absent) or absent from one or both chromosomes in some but not all individuals in a breed (Polymorphic).

Status	BSW	HOL	JER	RDC
Retained	76,320	73,101	75,499	71,650
Absent	163	177	680	14
Polymorphic	4,550	7,862	4,958	9,476
NA	108	1	4	1

Discussion

Retrotransposon-like elements as detected by *Ltrharvest* are common in the bovine genome. As has been observed in other species numerous putative retrotransposons are found more or less uniformly spread across the *Bos taurus* genome. Many of the putative retrotransposons were polymorphic in one or more or even all the breeds. This suggests extensive activity in mobilization of retrotransposons in the bovine genome.

The highest proportion of polymorphic putative retrotransposons was found in the RDC breed. This is consistent with generally high levels of polymorphism in this breed and its composition of several relatively distinct sub breeds and genetic contribution to the Danish Red Dairy breed including from BSW and HOL. Also, the RDC breed was missing only 14 of the putative retrotransposons observed in the Hereford genome. In contrast, the most distinctive breed, JER had the highest number (680) of putative retrotransposons missing.

The detection presented here is limited in that only retrotransposon-like elements present in the individual sequenced to produce the cattle genome assembly could be mined by *Ltrharvest*. The source of the data for the UMD 3.1 genome assembly was a Hereford bull (Zimin, Delcher, Florea et al. 2009), that is from an animal belonging to a beef breed. The animals whose genomes were analyzed here all belonged to dairy breeds. That means that most if not all the elements found to vary in the populations either have arisen since the separation of the ancestors of the Hereford or the ancestral population was polymorphic for the presence of the elements.

Another limiting factor in the approach taken here is that it is difficult to detect the complete absence of a putative retrotransposon. As retrotransposons are related detection of the “payload” of the retrotransposon in itself would not yield clear answers. The best way for large-scale detection of the presence of specific retrotransposons would probably be to search for read spanning the boundary between the LTR and the surrounding genomic sequence.

The presence of substantial amounts of polymorphism and/or the establishment of numerous new elements in the genome suggests that retrotransposon-like elements at present exhibit substantial amounts of mobility in the Bovine genome.

Conclusion

Putative retrotransposons are widespread across the bovine genome. Individual retrotransposons appear to vary in their presence/absence both within and between breeds.

References

- Adelson, D. L., Raison, J. M., and Edgar, R. C. (2009). *Proc. Natl. Acad. Sci., USA*, 106:12855–12860.
- Almeida, L. M., Silva, I. T., Silva Jr., W. A. et al. (2007). *Gene*, 390:180–189.
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). *BMC Bioinformatics*, 9:18.
- Janoušek, V., Karn, R., and Laukaitis, C. (2013). *BMC Evolutionary Biology*, 13:107.
- Kadri, N. K., Sahana, G., Charlier, C. et al. (2014). *PLoS Genetics*, 10:e1004049.
- Kazazian Jr, H. H. (1998). *Current Opinion in Genetics & Development*, 8:343–350.
- McClintock, B. (1956). *Cold Spring Harbor Symp. Quant. Biol.*, 21:197–216.
- Mills, R. E., Bennett, E. A., Iskow, R. C. et al. (2007). *Trends in Genetics*, 23:183–191.
- Smit, A. F. (1996). *Current Opinion in Genetics & Development*, 6:743–748.
- Zimin, A., Delcher, A., Florea, L. et al. (2009). *Genome Biology*, 10.