

POSTER PRESENTATION

Open Access

# Genes and genomes, an imperfect world: comparison of gene annotations of two *Bos taurus* draft assemblies

Liliana Florea<sup>1\*</sup>, Alexander Souvorov<sup>2</sup>, Steven L Salzberg<sup>1</sup>

From Beyond the Genome: The true gene count, human evolution and disease genomics  
Boston, MA, USA. 11-13 October 2010

## Background

Gene annotation is the first and most important step in analyzing a genome. As an increasing number of species are being sequenced and assembled to various degrees of completion, two key questions are: how does the quality of the assembled sequence affect gene annotation? What is the impact on scientists using the annotation?

## Results

We compared the gene annotations produced with the GNOMON [1] annotation pipeline for two *Bos taurus* genome assemblies produced at the University of Maryland [2,3]. We find that the changes to the assembly from one release to the next, which were quite substantial, made significant differences in the gene sets and gene structures, and implicitly the predicted proteins. The annotation was affected by the availability of new gene evidence and by seemingly rare genome mis-assembly events and local sequence variations. For instance, although the later assembly is generally superior, hundreds of protein coding genes in the earlier assembly are missing from the annotation of the later genome, and ~15% (~3600) of the genes have complex structural differences between the two assemblies. In addition, 15-20% of the predicted proteins have relatively large sequence differences when compared to their Ref-seq models.

## Conclusions

These findings highlight the consequences of genome quality on gene annotation, and argue for continued improvements in a genome sequence until it is truly

finished. They also demonstrate the challenges that confront a biologist when tracking a gene of interest between different assembly versions. In addition, these analyses have helped us identify specific loci for improvement, which will benefit the user community of the *Bos taurus* genome.

## Author details

<sup>1</sup>Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA. <sup>2</sup>National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA.

Published: 11 October 2010

## References

1. The NCBI GNOMON eukaryotic gene prediction tool. [<http://www.ncbi.nlm.nih.gov/genome/guide/gnomon.shtml>].
2. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassel CP, Sonstegard TS, et al: A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol* 2009, **10**:R42.
3. *Bos taurus* assembly site at the University of Maryland. [[http://www.cbcb.umd.edu/research/bos\\_taurus\\_assembly.shtml](http://www.cbcb.umd.edu/research/bos_taurus_assembly.shtml)].

doi:10.1186/gb-2010-11-S1-P13

Cite this article as: Florea et al.: Genes and genomes, an imperfect world: comparison of gene annotations of two *Bos taurus* draft assemblies. *Genome Biology* 2010 **11**(Suppl 1):P13.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



<sup>1</sup>Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA  
Full list of author information is available at the end of the article