A population scale discovery of large deletions in Holstein, Jersey and Nordic Red Cattle genome

Md Mesbah-Uddin[1,2], Bernt Guldbrandtsen[1], Terhi Iso-Touru[3], Johanna Vilkki[3], Dirk-Jan De Koning[4], Didier Boichard[2], Mogens Sandø Lund[1] and Goutam Sahana[1]

[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark.
[2]UMR 1313 GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
[3]Green Technology, Natural Resources Institute Finland, Myllytie 1, FI-31600, Jokioinen, Finland.
[4]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden.

Conclusions

-- We provide a high resolution genetic map of large deletions, which will be a valuable resource for gene mapping and genomic prediction in cattle

-- Deletion-breakpoints could be identified by targeted local assembly, verified by both known deletions and new discovery, which facilitates understanding of deletion formation mechanisms as well as custom chip design for routine genetic evaluation

-- The stratified deletions are interesting targets for investigating deletions undergoing genetic drift or artificial selection

Objectives

-- To map large deletions, and analyze their population-genetic properties in dairy cattle

-- To understand deletion formation mechanisms

Methods

-- Deletion discovery and genotyping, leveraging technical and population-level sequence features in 175 dairy cattle, using GenomeSTRiP-2 [1]

-- Validation using (1) 777K BovineHD chip intensity data, (2) breakpoint-assembly and alignment, and (3) PCR + Amplicon sequencing

Results

-- Around 8,500 large deletions (199bp to 773KB) with an overall FDR of 8.8%, 82% of them are novel compared to deletions in *dbVar* database

-- We identified breakpoints of a ~525 KB deletion on chromosome 23, causing stillbirth [2] in dairy cattle (Fig 1)

-- Majority of the deletions are generated by microhomology-mediated end joining

-- 158 highly differentiated deletions, ~27% of these overlap genic elements, e.g. exons, introns, or UTRs (Fig 2)

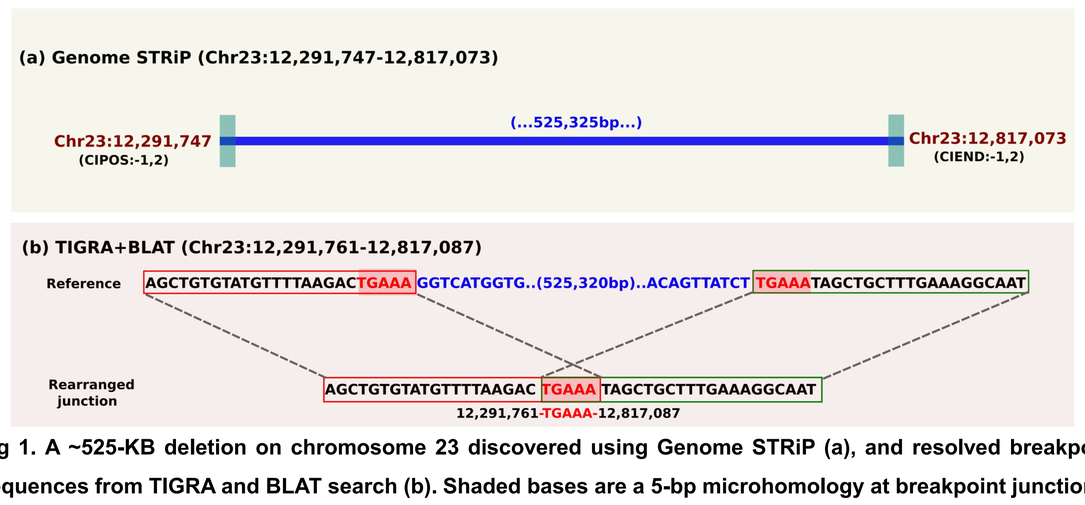

Fig 1. A ~525-KB deletion on chromosome 23 discovered using Genome STRiP (a), and resolved breakpoint sequences from TIGRA and BLAT search (b). Shaded bases are a 5-bp microhomology at breakpoint junction.

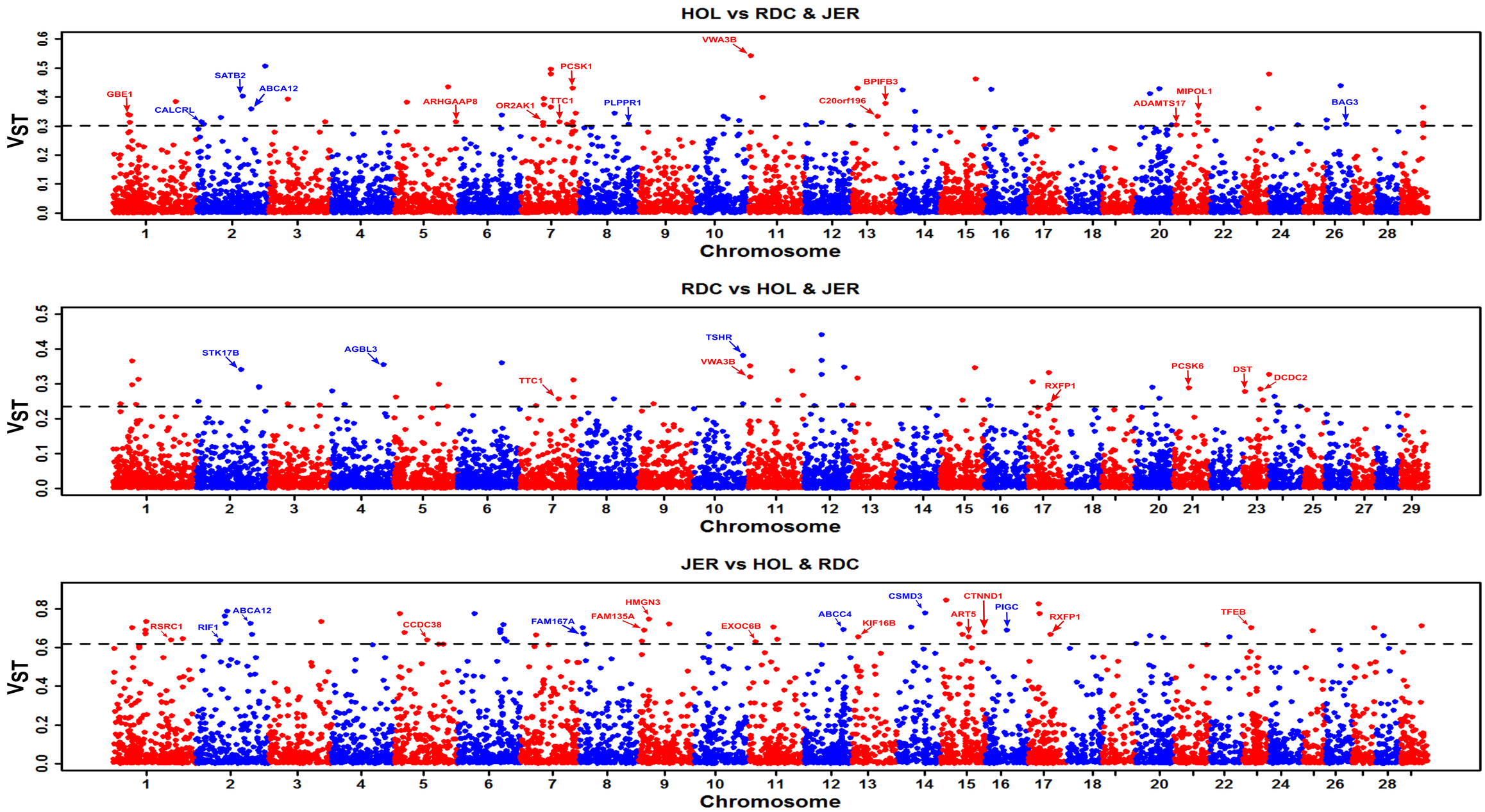

Fig 2. Population stratification based on V_{ST} (a measure of differentiation for structural variant, highly correlated with Wright's fixation index F_{ST}). Horizontal dash line indicates highly stratified deletion regions (V_{ST} ≥ Mean + 4 Standard deviation). HOL: Holstein; RDC: Nordic Red Cattle; JER: Jersey

mesbah@mbg.au.dk

Acknowledgements

Md Mesbah-Uddin benefited from a joint grant from the European Commission within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG". This research was supported by the Center for Genomic Selection in Animals and Plants (GenSAP) funded by Innovation Fund Denmark (grant 0603-00519B).

Reference

1. Handsaker RE, Korn JM, Nemesh J, McCarroll SA: **Discovery and genotyping of genome structural polymorphism by sequencing on a population scale.** Nat Genet 2011, 43(3):269-276.
2. Sahana G, Iso-Touru T, Wu X, Nielsen US, de Koning DJ, Lund MS, Vilkki J, Guldbrandtsen B: **A 0.5-Mbp deletion on bovine chromosome 23 is a strong candidate for stillbirth in Nordic Red cattle.** Genet Sel Evol 2016, 48:35.

A population scale discovery of large deletions in Holstein, Jersey and Nordic Red Cattle genome

Md Mesbah-Uddin[1,2], Bernt Guldbrandtsen[1], Terhi Iso-Touru[3], Johanna Vilkki[3], Dirk-Jan De Koning[4], Didier Boichard[2], Mogens Sandø Lund[1] and Goutam Sahana[1]

[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark.
[2]UMR 1313 GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
[3]Green Technology, Natural Resources Institute Finland, Myllytie 1, FI-31600, Jokioinen, Finland.
[4]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden.

Abstract

Structural variants (SVs) are DNA polymorphisms involving more than 50 base pairs, e.g. insertions, deletions, duplications and inversions, as well as more complex changes, have a wide-spectrum of phenotypic impacts ranging from beneficial to lethal in both humans and animals. Among SVs, large deletions are potential candidates for loss-of-function, which could be lethal as homozygotes when including essential genes. In this study, we scanned the whole-genome sequences (WGS) of 175 dairy cattle from three breeds (e.g. 67 Holstein, 27 Jersey, and 81 Red Dairy Cattle) to discover large deletions. We analyzed population genetic properties of these deletions, and explored their functional impact. We also explored the probable mechanisms of deletion formation for a set of breakpoint-resolved deletions. WGS reads were aligned to the bovine reference genome assembly *UMD3.1* using *BWA,* and deletions were detected and genotyped using *GenomeSTRiP-2.0*. We validated the results using three approaches: (i) using 777K BovineHD chip intensity data, (ii) breakpoint assembly and alignment by TIGRA and BLAT, and (iii) PCR + sequencing of amplicons. Here we report 8,480 large deletions (overall FDR 8.8%) − 82% of which are novel compared to deletions in the *dbVar* database. Breakpoint sequence analyses revealed that majority (24 of 29 tested) of the deletions contain microhomology/homology at breakpoint, and therefore, most likely generated by *microhomology-mediated end joining*. We observed higher differentiation among breeds for deletion variants in some genic regions, such as *ABCA12*, *TTC1*, *VWA3B*, TSHR, *DST/BPAG1*, and *CD1D*. The analysis of selective constraints (*dN/dS* of cow-mouse 1-to-1 orthologues) reveals that genes overlapping deletions are on average evolutionarily less conserved than known (mouse) lethal genes ($p\text{-value}=2.3\times10^{-6}$; *Wilcoxon* test). We also report 167 natural gene knockouts in cattle that are apparently nonessential as live homozygote individuals are observed. These genes are functionally enriched in immunoglobulin domains, olfactory receptors, and MHC classes (FDR=2.06×10^{-22}, 2.06×10^{-22}, 7.01×10^{-6}, respectively). We also demonstrate that deletions are enriched for health and fertility related QTL (2 and 1.5 fold enrichment, *Fisher's* $p\text{-value}=8.91\times10^{-10}$ and 7.4×10^{-11}, respectively). Finally, we confirmed and identified the breakpoint of a ~525 KB deletion on Chr23:12,291,761-12,817,087 (overlapping *BTBD9*, *GLO1* and *DNAH8*), previously reported to be associated with stillbirth in Nordic Red Cattle. This deletion catalog will facilitate discovery, genotyping, and imputation of deletions in large cohorts of animals, and subsequent studies for gene mapping and genomic prediction of breeding values.