P1147

Using an Online Genome Resource to Identify Myostatin Variation in U.S. Sheep

M.P. Heaton[1], T.P.L. Smith[1], and T.S. Kalbfleisch[2],

1) USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, USA, 2) University of Louisville, Louisville, KY, USA.

Abstract

We created a public, searchable DNA sequence resource for sheep that contained approximately 16x whole genome sequence of 96 rams. The animals represent 10 popular U.S. breeds and share minimal pedigree relationships, making the resource suitable for viewing gene variants in the user-friendly Integrated Genome Viewer (IGV) environment. To illustrate its use, the DNA sequence reads were viewed for myostatin, a gene encoding a negative regulator of skeletal muscle growth. Two putative functional variants were observed, both of which had been reported previously. One variant creates a binding site for a miRNA in the 3'UTR that reduces the abundance of myostatin protein. The other variant changes a glutamate (E) residue to glycine (G) at position 34. By viewing these variants in IGV, it was simple to estimate their frequencies in these 96 rams. The 3'UTR variant allele was homozygous in 9 of 10 Texel rams, while G34 carriers were present in Dorset, Navajo Churro, Rambouillet, and USMARC composite breeds. In addition, one Dorset ram was homozygous for the G34 allele. The strict evolutionary conservation of the E34 allele throughout the Amniota clade of tetrapods, combined with the multi-breed distribution of the putative reduced function G34 allele in sheep, is consistent with the hypothesis that the G34 allele could interfere with myostatin function and positively affect muscle growth in U.S. sheep. This study provides a new resource for discovering potentially functional variants, and making initial rapid *in silico* estimates of allele frequency among U.S. breeds.

Viewing, depth, and accuracy

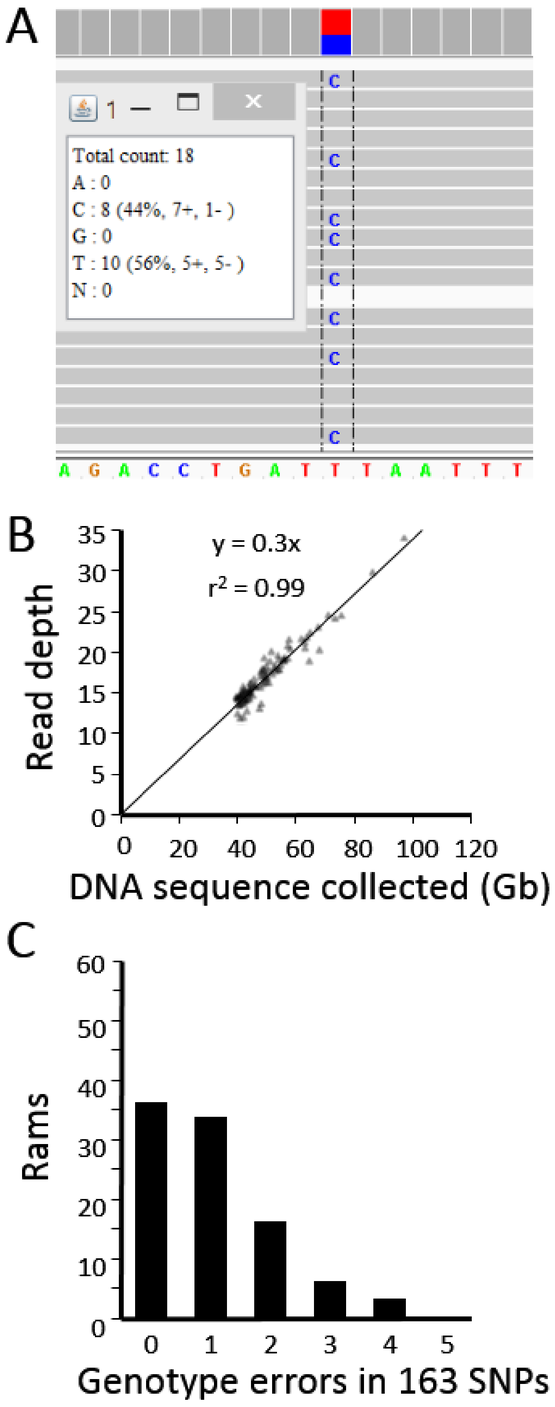

Figure 2. Comparison of 163 reference (parentage) SNP genotypes with those derived from WGS data. Panel A, computer screen image of one animal's WGS data aligned to ovine reference assembly OAR3.1 at a parentage SNP site. The heterozygous C/T genotype is shown as viewed with the Integrated Genome Viewer (IGV) software. Panel B, linear relationship between mapped read depth and the amount (Gb) of Q20 WGS data collected. At each SNP position, the read depth and genotypes were visualized and manually recorded for 163 parentage SNPs. Panel C, genotype scoring accuracy for 163 parentage SNPs in the 96 rams. Consensus reference genotypes (n = 15,648) for the parentage SNPs were previously determined by multiple independent methods.

Breeds

Figure 1. Breeds represented in the USMARC Sheep Diversity Panel version 2.4. To maximize the total number of unshared haploid genomes, rams were selected for pedigrees with minimal relationships.

Coverage & concordance

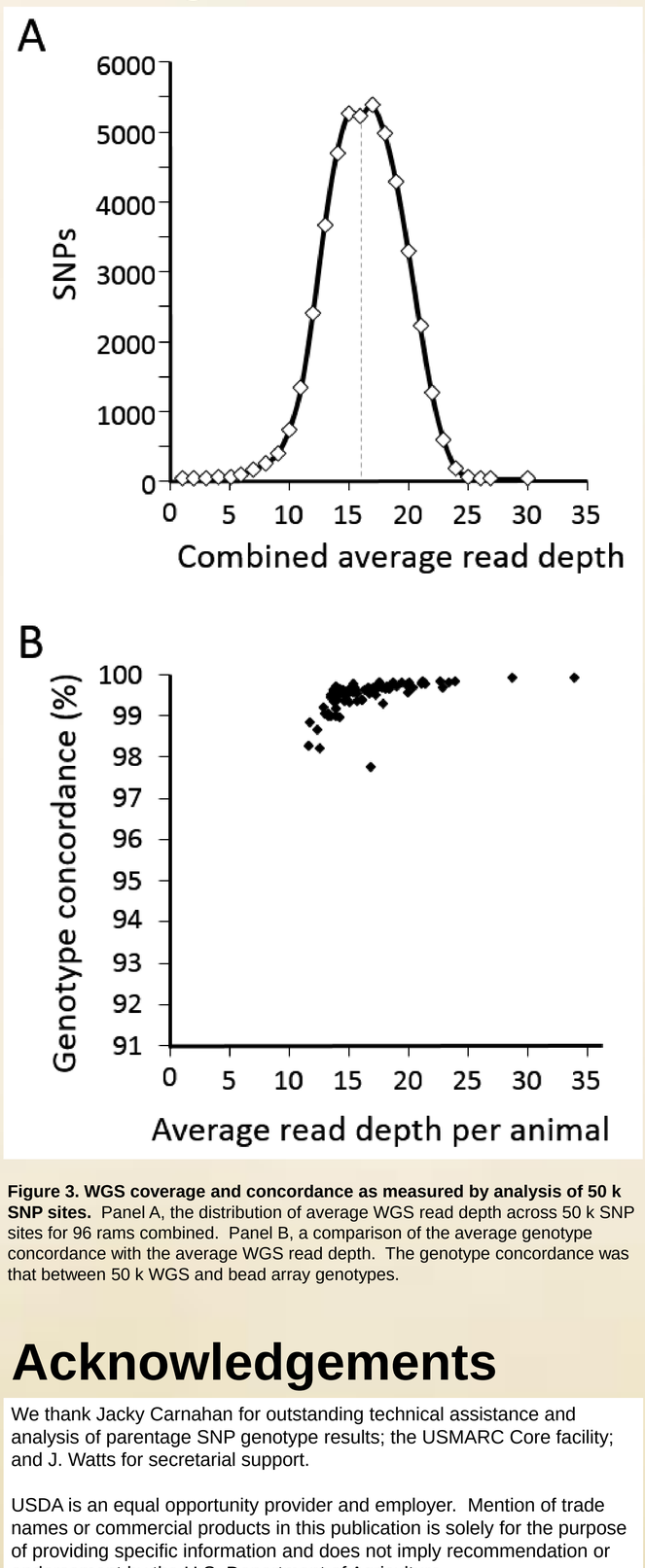

Figure 3. WGS coverage and concordance as measured by analysis of 50 k SNP sites. Panel A, the distribution of average WGS read depth across 50 k SNP sites for 96 rams combined. Panel B, a comparison of the average genotype concordance with the average WGS read depth. The genotype concordance was that between 50 k WGS and bead array genotypes.

Acknowledgements

We thank Jacky Carnahan for outstanding technical assistance and analysis of parentage SNP genotype results; the USMARC Core facility; and J. Watts for secretarial support.

USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Agricultural Research Service

USMARC

UNIVERSITY OF LOUISVILLE

Outcomes

- Average coverage 16x for 96 diverse rams
- Accuracy and call rate > 99% for WGS
- Publicly viewable with IGV and files at NCBI
- E34G coding variant found in U.S. sheep

Conclusion

Three hours of public IGV viewing identified variants that previously required an intensive PCR/Sanger sequencing effort to complete.

MSTN E34G variant

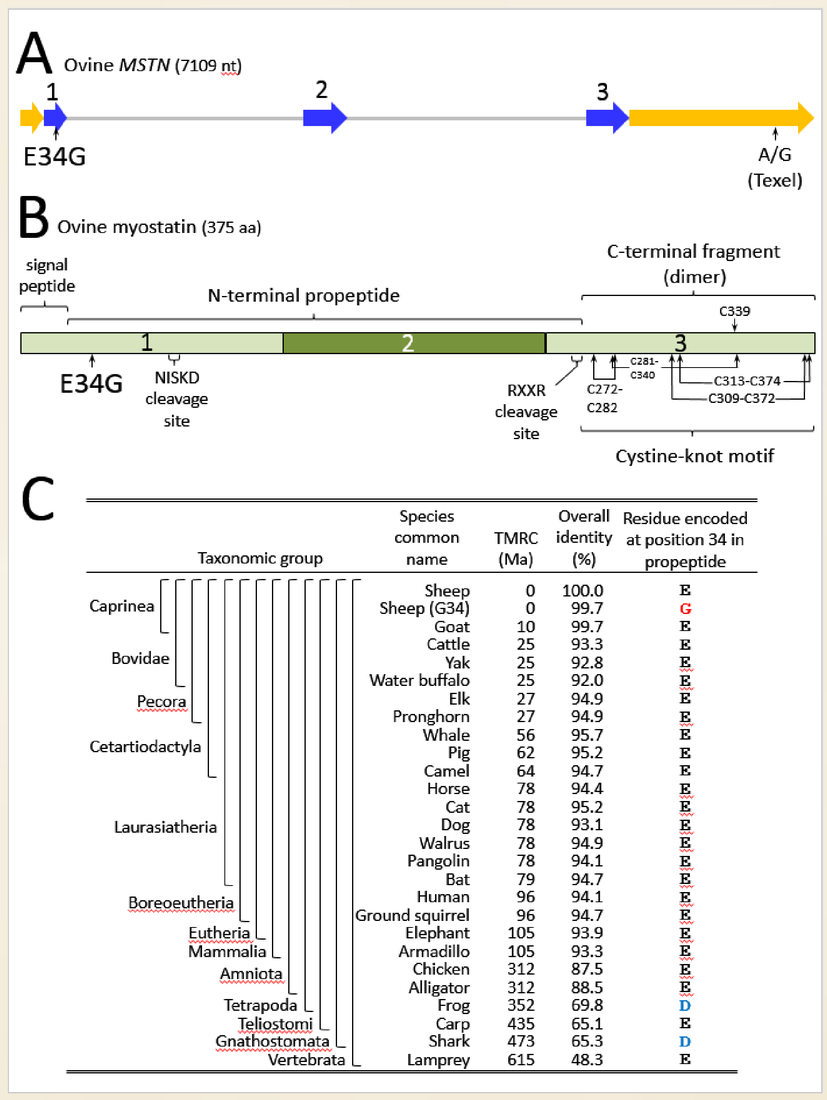

Figure 4. Physical maps and evolutionary comparison of the myostatin E34G variant found in US Sheep. Panel A, genomic DNA map of MSTN: orange arrows, UTR regions; blue arrows, coding sequence regions; grey horizontal lines, intron regions. Panel B, map of myostatin protein domains. Panel C, Evolutionary comparison of myostatin E34G coding variant site. Abbreviations: TMRC (Ma), is the estimated time to most recent common ancestor in millions of years; D, aspartic acid.

How to find the online resource