Impact of Pedigree Information and Genome Assembly Errors on Inference of SNP Haplotypes in Cattle

A.S. Leaflet R2770

Ziqing Weng, Graduate Student; Dorian Garrick, Professor, Department of Animal Science; Mahdi Saatchi, Postdoctoral Research Associate, Iowa State University;
Robert Schnabel, Research Associated Professor;
Jeremy Taylor, Professor, Division of Animal Science, University of Missouri

Summary and Implications

The impact of pedigree information and SNP location determined from either the UMD3.1 genome sequence assembly or the USDA-AIPL map on phasing accuracy were evaluated for 2 chromosomes in 2,778 parent verified Angus sire/offspring pairs. DAGPHASE (Druet and Georges, 2010), using a single generation pedigree was superior to BEAGLE software (Browning and Browning, 2007) for phasing. Results based on USDA-AIPL map are closer to expectation than those based on UMD3.1, but the difference is not significant. Recombination hotspots were detected near 4 and 82Mb on BTA14, and near 25Mb on BTA15.

Introduction

The accuracy of genotype imputation depends on the accuracy of phasing haplotypes from single nucleotide polymorphism (SNP) genotypes. One approach to judge phasing accuracy is to estimate the number of recombination events by comparing diplotypes within parent/offspring pairs. This study evaluated the impact of using pedigree information and the impact of alternative maps of SNP locations on phasing accuracy, by quantifying recombination events in meioses between sire and offspring.

Materials and Methods

Two thousand seven hundred and seventy-eight parent verified Angus sire/offspring pairs with BovineSNP50 genotypes were used in this study. SNPs were removed with call rate <0.95, minor allele frequency <0.01, p value for a Hardy Weinberg Equilibrium test <0.001, or Mendelian inconsistency rate >0.0024 (95% quantile). Phasing of Bos taurus (BTA) chromosomes 14 and 15 was performed using the unrelated option within BEAGLE or using the pedigree-based DAGPHASE with the UMD3.1 or USDA-AIPL maps of SNP order.

Results and Discussion

The expected numbers of recombination events was estimated using the Haldane map function, and for BTA14 were (#cattle x #recombination) 1187x0, 1009x1, 429x2, and 153x>2; and for BTA15 were 1176x0, 1011x1, 435x2, and 156x>2. The observed recombination events on each chromosome using different phasing software and maps are reported in table 1. Average numbers of cattle observed with >2 recombination events were 928 using BEAGLE and 467 using DAGPHASE. Both methods overestimated recombination events, but less so with DAGPHASE than BEAGLE. The average number of cattle with >2 recombination events estimated using DAGPHASE and UMD3.1 locus ordering was 544, but was 417 using the USDA-AIPL map. Recombination hotspots were detected near 4 and 82Mb on BTA14, and near 25Mb on BTA15. Possible reasons for overestimation of recombination are inadequate half sib family sizes in this data, and SNP location errors on the linkage maps.

Acknowledgments

We gratefully acknowledge the constructive suggestions from our colleagues in Iowa State University.

Table 1. The observed recombination events on each chromosome

Method	Map_BTA	# of observed recombination			
		# =0	#=1	#=2	#>2
DAGPHASE	UMD3_14	677	993	577	531
	USDA_14	787	1100	487	404
	UMD3_15	670	957	593	558
	USDA_15	757	1060	531	430
BEAGLE	UMD3_14	575	763	519	928
	USDA_14	703	884	361	686
	UMD3_15	530	690	528	924
	USDA_15	706	919	398	644