White Paper

Development of a high-throughput, high-density bovine genotyping array

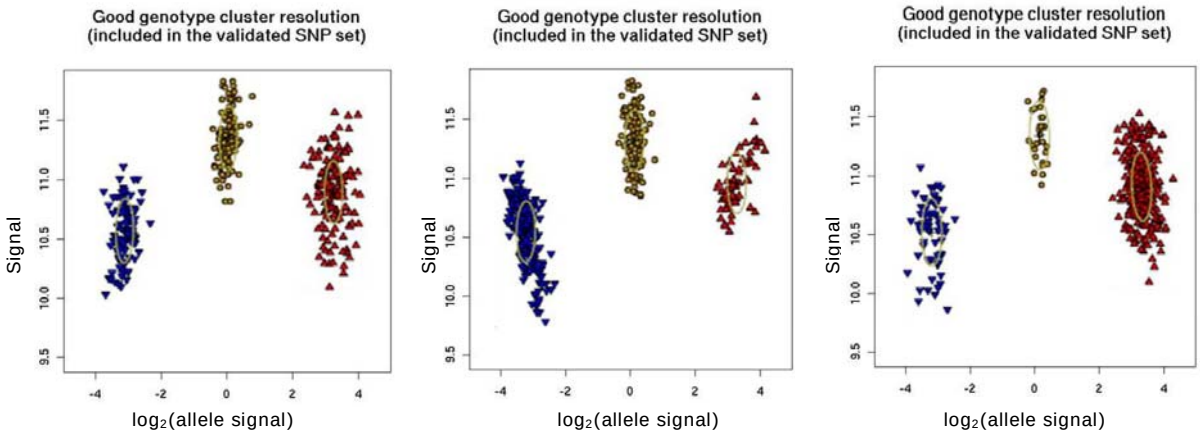
Background

Simultaneous genotyping of many single nucleotide polymorphisms (SNPs) has been made possible by the development of array-based hybridization platforms. High-density genotyping empowered the success of genome-wide association studies for determining genetic variation affecting complex traits. Medium-density studies in agricultural organisms have been beneficial in marker-assisted breeding, mapping quantitative trait loci, and other applications. Now there is interest in expanding to higher-density options to further refine and develop these techniques. The bovine research community has undertaken a massive genome sequencing and SNP screening effort to develop a comprehensive solution for a versatile, high-throughput, high-density bovine SNP genotyping array.

SNP screens

On the Axiom™ Genotyping Solution from Affymetrix, SNP screening was carried out for the SNPs identified by the sequencing effort. For SNPs to be considered validated, they must be robust, measured by performance metrics such as genotype call rate, cluster separation, reproducibility, etc., and they must exhibit at least two examples of the minor allele (Figure 1).

Figure 1: Cluster separation in the $\log_2$(allele signal) difference space is one of the metrics applied in the SNP screens. Each sample is shaped and colored according to called genotypes (red, yellow, blue, and gray correspond to AA, AB, BB, and no call, respectively). Ovals are cluster position posterior results from the genotyping algorithm.



Experimental design

The Affymetrix® Bovine Consortium, consisting of academic researchers and breeding groups, has combined efforts to sequence genomes of 15 bovine breeds from *Bos indicus* and *Bos taurus* on three next-generation sequencing platforms to produce an extensive collection of SNPs and their genotypes. Screening arrays were designed using a method to optimize physical coverage of a subset of these SNPs across the sequenced genomes. A diverse set of 384 samples, representing taurine, indicine, tropical taurine, and Asian breeds, was obtained from the HapMap collection and other collaborators and screened on the screening arrays on the high-throughput Axiom Genotyping Solution.

Results

An analysis of 3 million validated, high-performance SNPs polymorphic across 20 breeds is presented (Tables 1 and 2). Between 0.7 million and 2.4 million SNPs were found to be polymorphic in each breed (Figure 2), and 31,000 validated exonic SNPs were identified (Figure 3). Principal component analysis (PCA) shows that breeds segregate into their major geographic types, with well-defined substructure within each type (Figure 4).

Table 1: Performance over 3 million SNPs. To measure the performance of more than 20 diverse breeds, the samples were split into training and test sets. Test set performance is shown.

	Test set
Average sample call rate	99.5%
Mendelian consistency (23 trios)	99.9%
Reproducibility (10 replicates)	99.8%
Gender calling accuracy	99.8%
Mean gap size	925 bases
Median gap size	502 bases

Table 2: Breed information.

Breed	Polymorphic SNPs	Avg. MAF*	Samples genotyped	Unrelated samples	Avg. call rate (%)
Afrikander	1.4M	0.27	5	4	99.5
Angus	1.4M	0.21	50	25	99.5
Ayrshire	0.96M	0.29	6	3	99.6
Blonde d'Aquitaine	0.99M	0.27	5	4	99.5
Boran	2.2M	0.26	10	9	99.5
Brahman	2.4M	0.23	23	19	99.1
Brown Swiss	0.99M	0.28	10	5	99.2
Simmental	1.4M	0.21	16	16	99.5
Gir	2.1M	0.24	20	15	99.3
Hanwoo	1.3M	0.25	11	5	99.4
Hereford	1.2M	0.22	21	15	99.5
Holstein	1.6M	0.21	56	34	99.5
Japanese Black	1.1M	0.24	10	8	99.6
Jersey	1.2M	0.23	29	13	99.5
Limousin	1.4M	0.21	20	17	99.5
Nelore	2.3M	0.22	50	38	99.4
Norwegian Red	1.0M	0.39	7	6	99.5
Rouge des Prés	0.7M	0.33	5	2	99.6
Romagnola	1.6M	0.20	26	18	99.6
Tuli	1.4M	0.29	4	3	99.5

*Average minor allele frequency for unrelated samples.

Figure 2: Number of validated SNPs found to be polymorphic in each breed. *Bos indicus* breeds are in green.

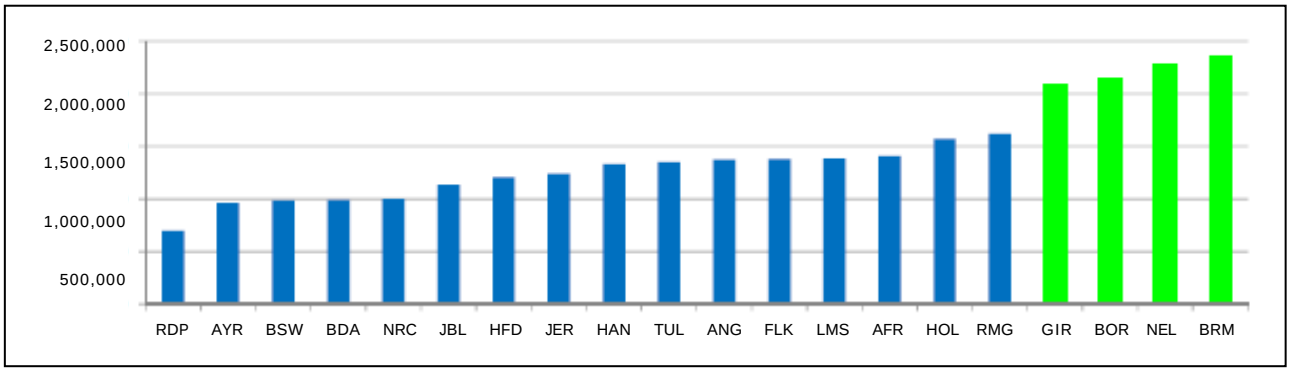


Figure 3: Validated exonic SNPs.

Biological category	No. of SNPs
Exon CDS SNPs nonsynonymous	8K
Exon CDS SNPs synonymous	12K
Exon UTR SNPs	8K
Exon other SNPs	3K
Splicing site SNPs	55
Intron SNPs	0.94M
Other SNPs	2M

Note: CDS = coding sequences; UTR = untranslated regions

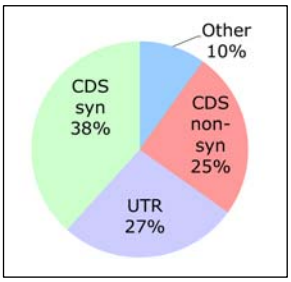


Figure 4A & 4B: PCA shows the breeds segregate based on subpopulations. Cladogram from Decker J. E., *et al*., *PNAS* **106**:18644-49 (2009).

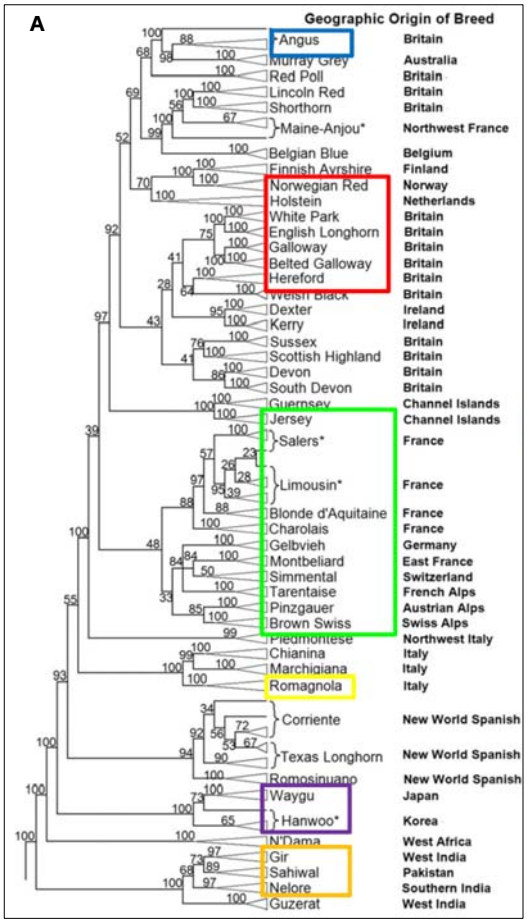
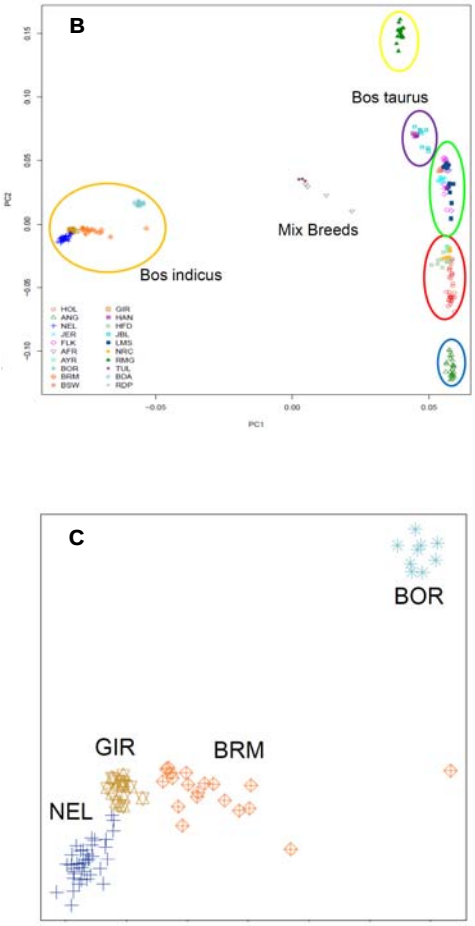


Figure 4C: PCA subplot showing novel resolution of *Bos indicus* breeds.



Conclusion

Affymetrix has developed a comprehensive collection of SNPs that genetically cover the linkage disequilibrium (LD) blocks of multiple cattle breeds. This SNP collection can be utilized for customized breed-specific array development. Future breed-specific screens will add to this pool of SNPs with validated genotyping performance. These high-density genotyping assays will support marker-assisted breeding and other applications in cattle.

©2011-12 Affymetrix, Inc. All rights reserved.