16750. Linkage disequilibrium in brown Swiss cattle with Illumina 50K and 777K arrays

Ehud Lipin[1], Alessandro Bagnato[2], Marlies Dolezal[2], Santos, E.[2], and Morris Soller[1]

[1]The Hebrew University of Jerusalem, 91904 Jerusalem, Israel; [2]Universita` degli Studi di Milano, 20133 Milano, Italy

Introduction

The purpose of this study was to compare bovine LD and haplotype block (HB) structure as it appears through the lens of the 50K and 777K Illumina bovine arrays, and to infer the potential gain of a 2.2M array.

Arrays and Samples

Italian Brown Swiss genomic Samples:

50K: 816 haplotypes; 777K: 384 genotypes.

Table 1. Array characteristics after Quality Control.

Array Characteristic	50K	777K
Nominal No. of SNPs on the array	51,582	735,237
Effective No. of polymorphic SNPs after QC	31,879	487,080
Effective marker density: SNPs/cM	3.14	162.09
Effective marker spacing: Kb/SNP	78.70	5.15

Non-Syntenic LD

- Non-syntenic LD was used to obtain low stringency base point for statistical significance.
  - Random non-syntenic and non-redundant pairs from different chromosomes were identified.
  - These pairs are by definition independent, representing LD under H0 (no linkage).
- LD is highly sensitive to sample size.
  - ⇨ S100. A random sample of 100 haplotypes to estimate effect of sample size.

Table 2.. Distribution of 50K non-syntenic LD values for the 50K array. N, number of LD values in the bin; Cum, cumulative proportion; All, all 816 HTs; S100, random sample of 100 HTs. Critical r^2, r^2 value exceeded by no more than 0.001 of non-syntenic marker pairs.

	All			S100		
r^2	N	Prop.	Cum	N	Prop.	Cum
≤ 0.01	2,937	0.979	0.979	1,979	0.661	0.661
>0.01 - =0.02	60	0.020	0.999	486	0.162	0.823
>0.02 - =0.03	3	0.001	1.000	253	0.085	0.908
>0.03 - =0.04				127	0.042	0.950
>0.04 - =0.05				73	0.024	0.974
>0.06 - =0.10				71	0.024	0.998
>0.11 - =0.15				5	0.002	0.9997
>0.16 - =0.20				1	0.0003	1.0000
Critical r^2	0.020			0.115		

Distribution of non-syntenic LD broader for S100, but does not exceed 0.15.

Haplotype blocks (HB)

- A stretch of consecutive markers with LD between adjacent pairs ≥ critical value.
- Stringency:
  - Low: LD≥0.15. Useful for Genomic selection.
  - High: LD≥0.70. Useful for GWAS.

Table 3. Average block size in bp. by number of marker pairs, stringency and array.

- HB size is measured in SNP pairs.
- Note: physical size (Kb) of the same "Pair – size" HB much smaller for 777K array than for the 50K array.

	50K		777K	
Pairs	LD≥0.15	LD≥0.70	LD≥0.15	LD≥0.70
1	65,672	46,953	5,759	4,555
2	134,047	98,301	11,225	8,466
3	211,353	153,070	16,019	11,804
4	262,502	197,222	20,981	15,018
5	334,743	283,201	24,765	17,799
6	429,024	322,105	29,608	19,773
7-10	541,595	338,160	38,768	26,143
11-20	936,232		62,847	41,587
21-30	1,537,037		101,629	73,012
31-40			144,754	115,076
41-50			190,401	144,391
51-60			220,133	209,492
61-70			251,299	335,701
71-80			320,159	247,713
81-90			427,003	528,719
91-100			447,434	838,651
>100			560,689	303,361

Acknowledgement. This study is funded by the EC 7th Framework Programme (FP7/2007-2013), agreement n° 222664, "Quantomics" (this Publication reflects only the author's views and the EC is not liable for any use that may be made of the information contained herein).

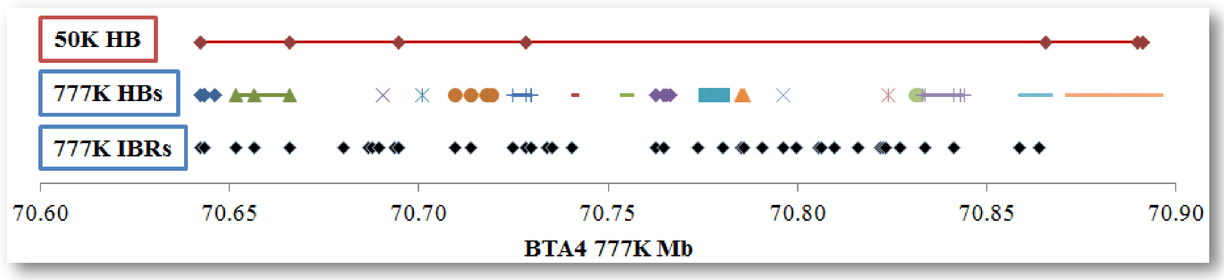
Table 4. Diagonal HB according to stringency and array, including only polymorphic SNPs after QC.

	LD≥0.15		LD≥0.70	
	50K	777K	50K	777K
Total HB	7,078	53,949	2,769	83,949
Total SNPs in HB	22,484	456,265	6,239	338,190
SNPs in HB/Total SNPs	0.705	0.937	**0.196**	**0.694**
Avg. marker pairs/HB	2.18	7.46	1.25	3.03
Avg SNPs/HB	3.18	8.46	2.25	4.03
Avg. Kb/HB	147.4	33.9	60.1	11.1
Max SNPs/HB	23	209	8	132

- At low stringency, the proportion of the genome captured in HBs does not differ greatly between the 50K array (70.5%) and the 777K array (93.7%). But at high stringency only a small fraction of the genome is captured in HBs by the 50K array (19.6%) compared to the 777K array (69.4%).
- The 777K array uncovers many more blocks compared to the 50K array, but the blocks are much smaller when measured in Kb.

⇨ **This implies that the HB uncovered by the 50K array are actually fragmented into composites made up of multiple 777K HBs.**

Figure 1. A 50K block is fragmented by the 777K array. A single high stringency 50K HB on BTA4 made of 7 marker pairs (= 8 SNPs) distributed over 396,613 bp, is found by the 777K to be fragmented, with no less than 17 high stringency sub-HBs made of 1-12 pairs (2-13 SNPs) distributed over 1,133-28,288 bp. Each 777K sub-HB has different color. haplotype block; IBR, inter-block region (marker pairs with LD<0.70).



2.2M array?

- There are plans for a 2.2M bovine array.

⇨ We looked at LD among 777K marker pairs separated by the average separation distance in a 2.2M array.

✓ 3.05 Gb sequence/2.2M SNPs = 1,386 bp/SNP.

Table 5. Distribution of diagonal LD against separation distance.

	bps		
r^2	≤500	501-1000	All
≤0.05	0.053	0.059	0.123
>0.1 - =0.2	0.045	0.053	0.087
>0.2 - =0.3	0.045	0.046	0.069
>0.3 - =0.4	0.040	0.039	0.058
>0.4 - =0.5	0.032	0.037	0.049
>0.5 - =0.6	0.037	0.035	0.045
>0.6 - =0.7	0.029	0.034	0.042
>0.7 - =0.8	0.033	0.034	0.043
>0.8 - =0.9	0.041	0.043	0.048
>0.9 - =1.0	0.646	0.620	0.435
>0.7	**0.720**	**0.696**	**0.527**

→ Marginal gains

- 52K → 777K: proportion of markers pairs in high LD (>0.7) increased 350%, 0.196 → 0.694 (Table 5).
- 777K → 2.2M (estimated): proportion of marker pairs in LD>0.7 increased by about 34%, 0.527 → 0.696-0.720 (average 0.708).

✓ There appears to be a limit at about 0.70 or 0.75 to the proportion of genome that can achieve LD>0.7.

– What determines this limit?

Table 6. Haplotype distribution for marker pairs as a function of r^2. HTs, haplotypes; HTs freq., average frequency of 10 random pairs.

		r^2			
		0.5	0.7	0.8	1
HTs Freq.	AB	0.595	0.708	0.660	0.693
	ab	0.274	0.229	0.295	0.307
	aB	0.125	0.051	0.038	0.000
	Ab	0.006	0.012	0.007	0.000
	AB+ab	0.869	0.937	0.955	1.000
	Ab+aB	0.131	0.063	0.045	0.000

- LD≥0.5 requires a very special combination of haplotypes, namely:
  - ✓ Two alternative haplotypes (AB, ab) athigh to moderate frequency,
  - \+ A third "recombinant" haplotype (aB) at low frequency,
  - \+ A fourth "recombinant" haplotype (Ab) at very low frequency.
- **Yet over 75% of diagonal pairs present LD>0.5.**
  - – How did this very special haplotype distribution come to occupy over 75% of the diagonal marker pairs?!
- ***For an explanation see Poster 16754: Marker to marker LD in relation to the evolutionary history of the site.***

האוניברסיטה העברית בירושלים
The Hebrew University of Jerusalem

UNIVERSITÀ DEGLI STUDI DI MILANO