

LD analyses in the sheep HapMap project using an ovine 50K SNP array

International Sheep Genomics Consortium
Sheep HapMap project



Herman Raadsma
on behalf of LD work team

With particular input

Dr Mehar Khatkar and Dr Ben Hayes

www.sheephapmap.org

Linkage Disequilibrium-LD

LD- non-random association of alleles at two or more loci,

LD used as input for

Population structure
Population diversity
Effective population size
Breed development

Genome Structure
Haplotype construction
Haplotype diversity
Selective sweeps

Building LD maps-
marker ordering

Usual indicators of LD

- r^2 range 0-1

- D' range 0-1

Resources

73 breeds - range

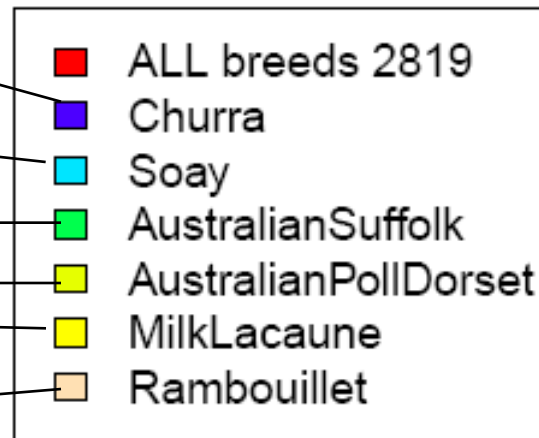
2819 individuals

50K genotype array- 49,034 genotypes post QC

Adjusted for sample size

Breeds n<20 excluded

6 breeds n>100



r^2 corrected for sample size

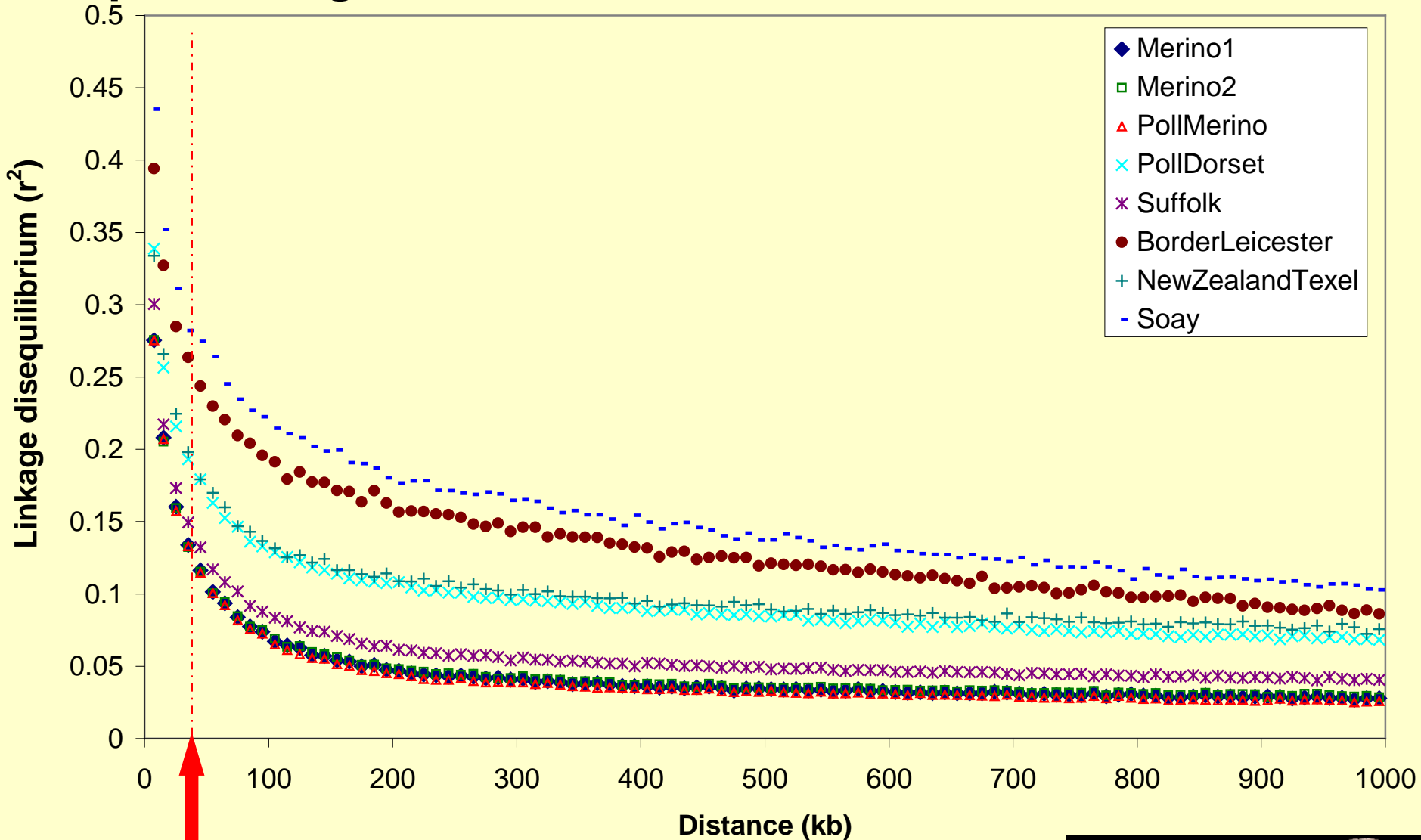
$$r^2(\text{corrected}) = (r^2 - 1/n) / (1 - 1/n),$$

where n is the number of haplotypes

(eg 25 animals would be 50 haplotypes).

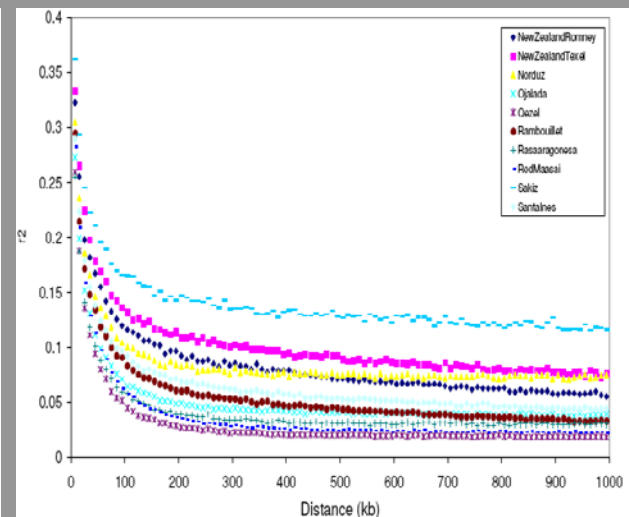
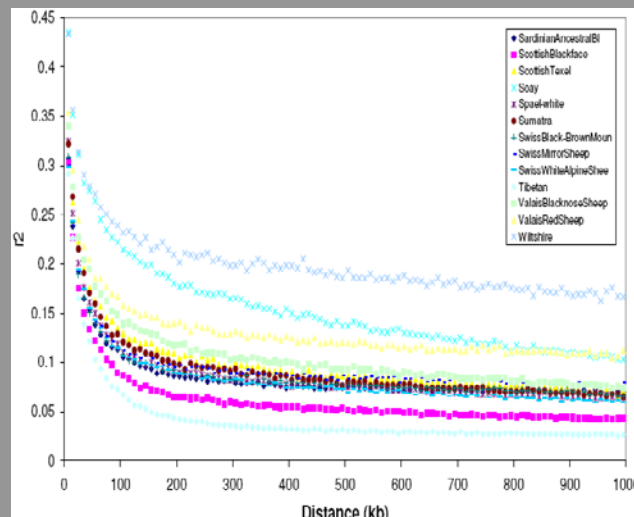
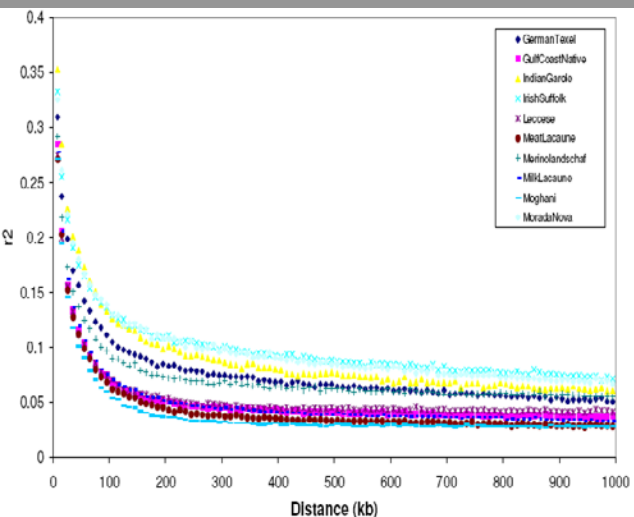
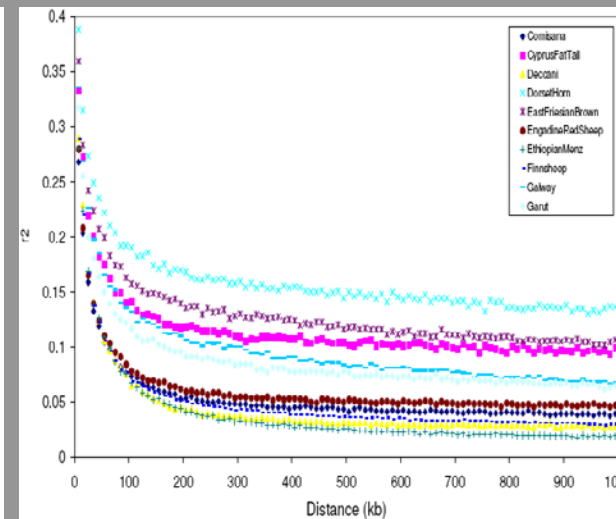
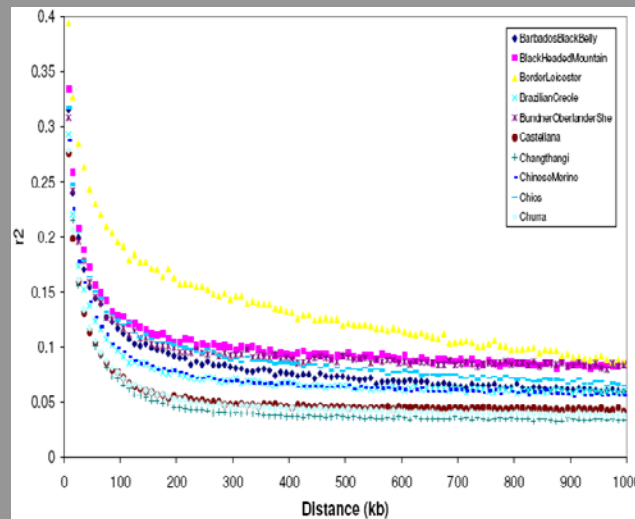
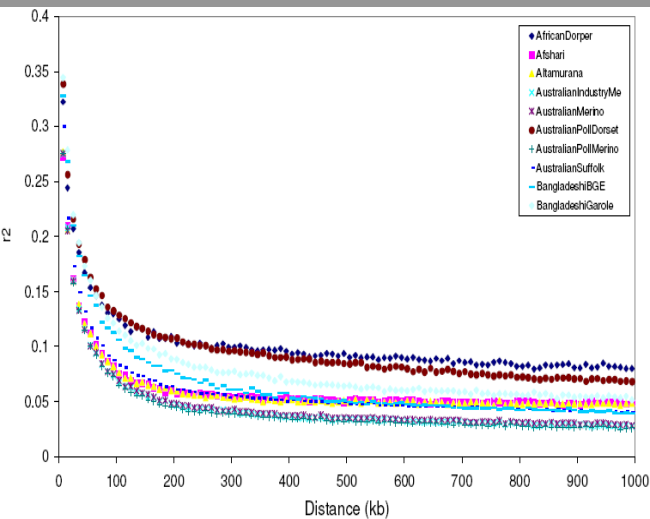
Range of LD profiles among 73 breeds- selected sample

Representing extremes



Average SNP spacing 50K

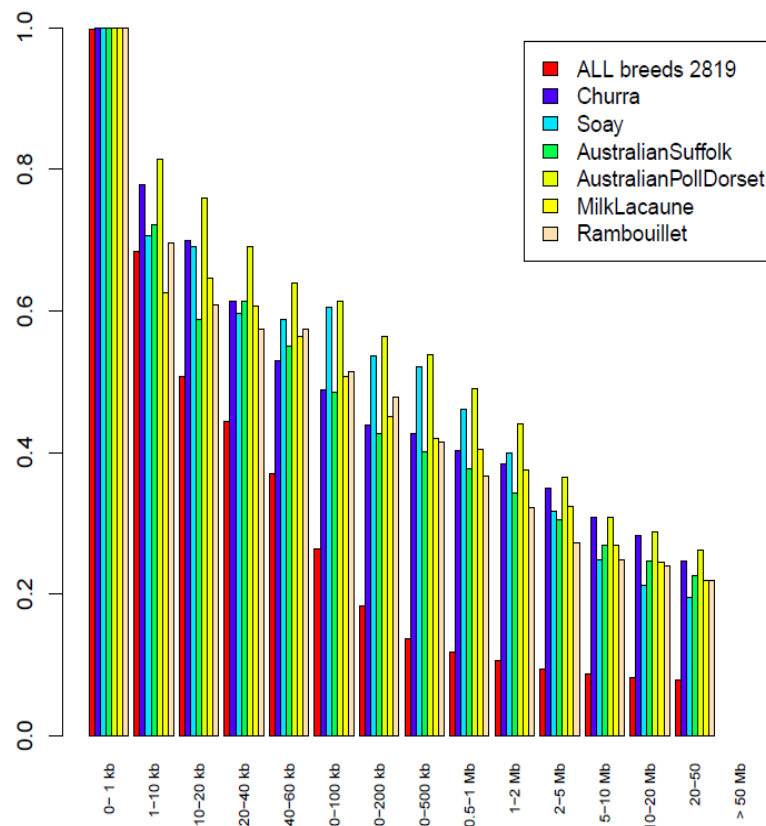
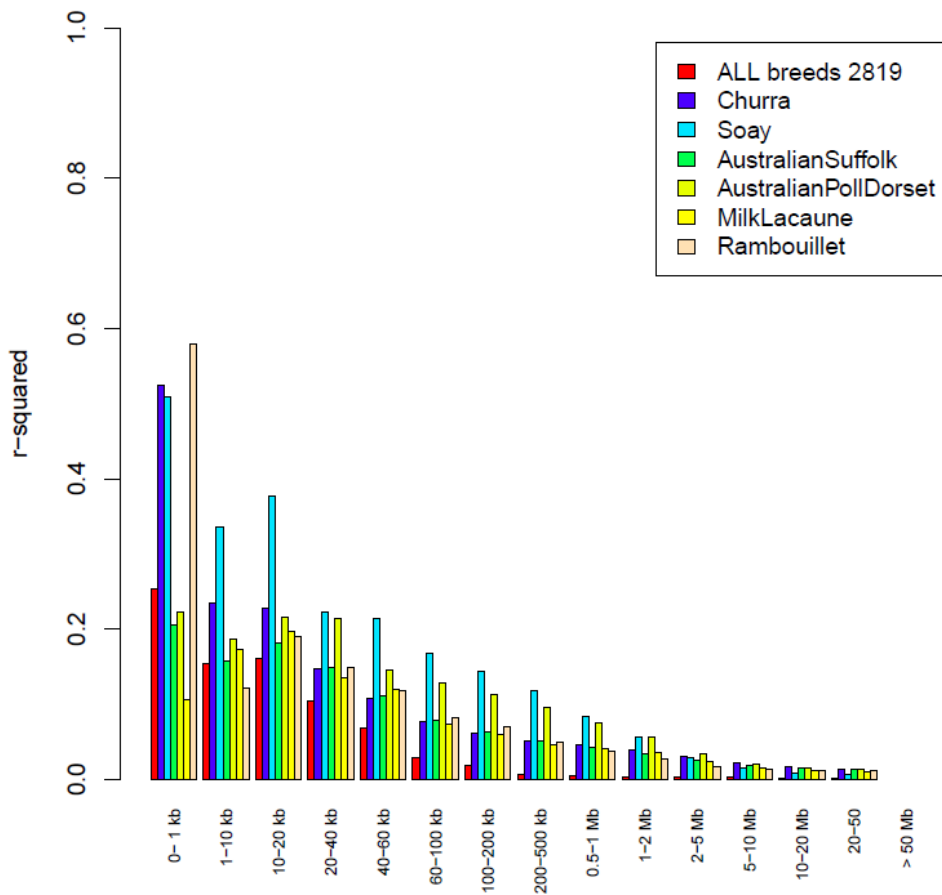
LD profiles all breeds



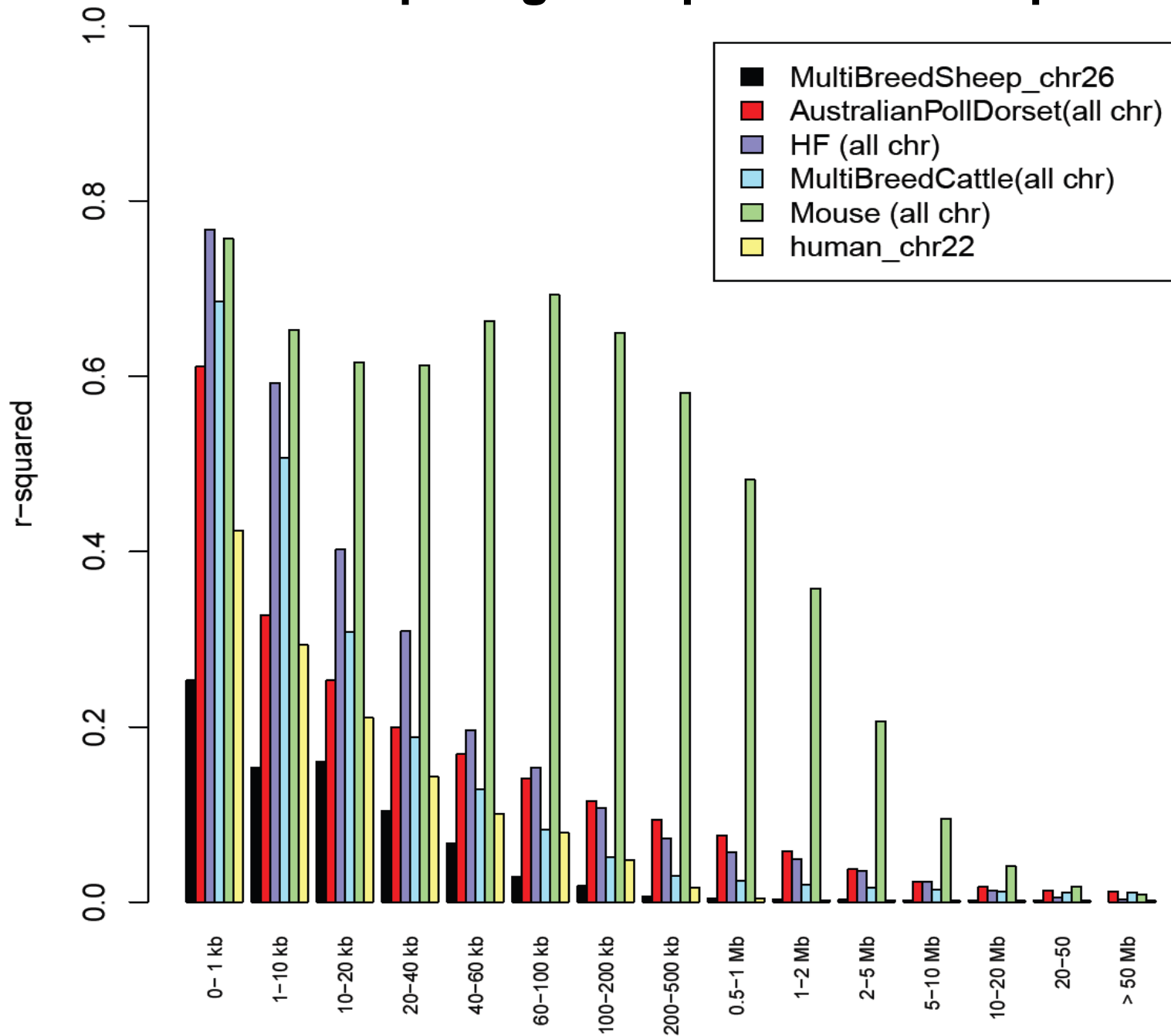
Comparison of breeds for LD for chromosome 26

r^2

D'



Comparing Sheep LD to other species

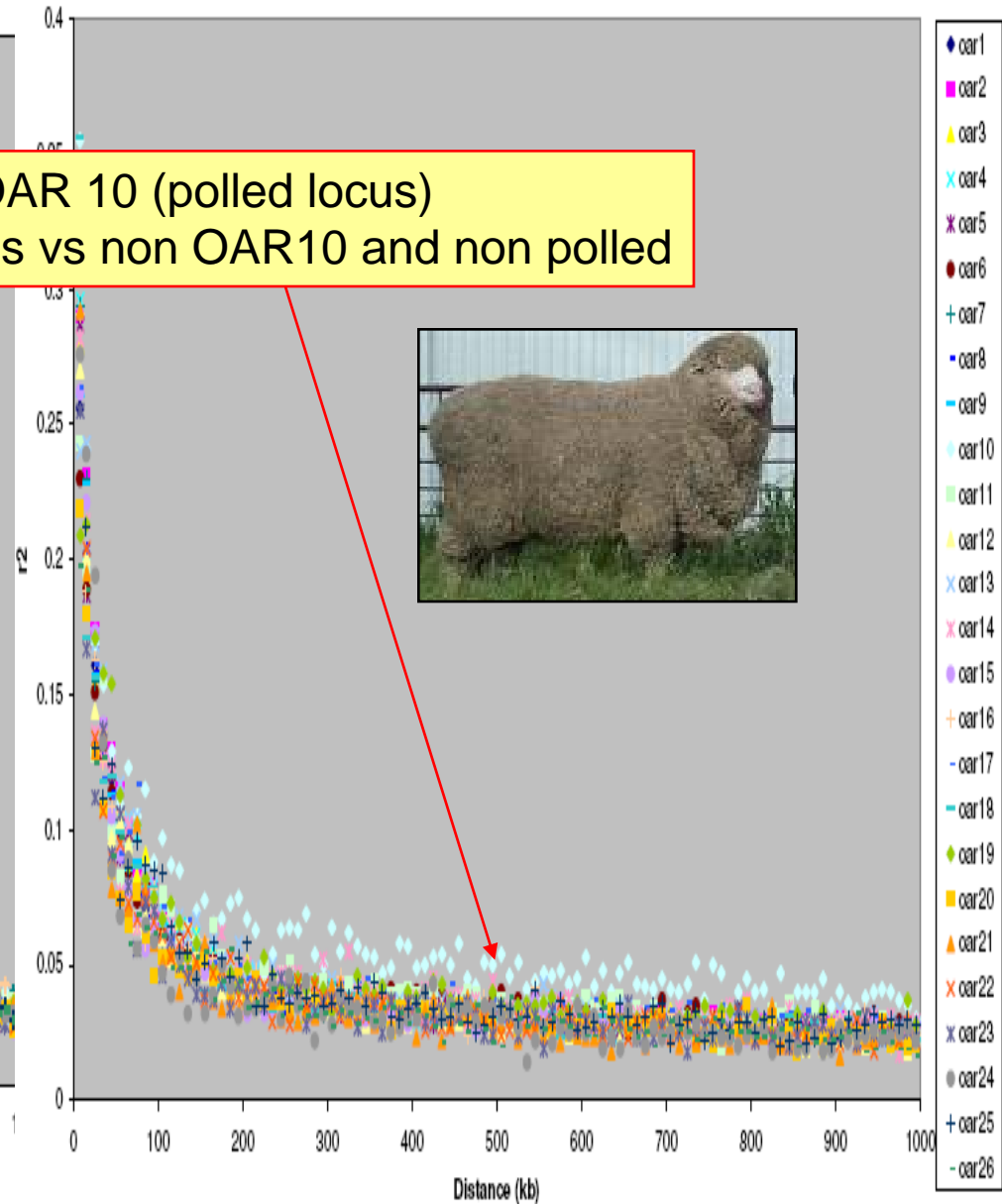
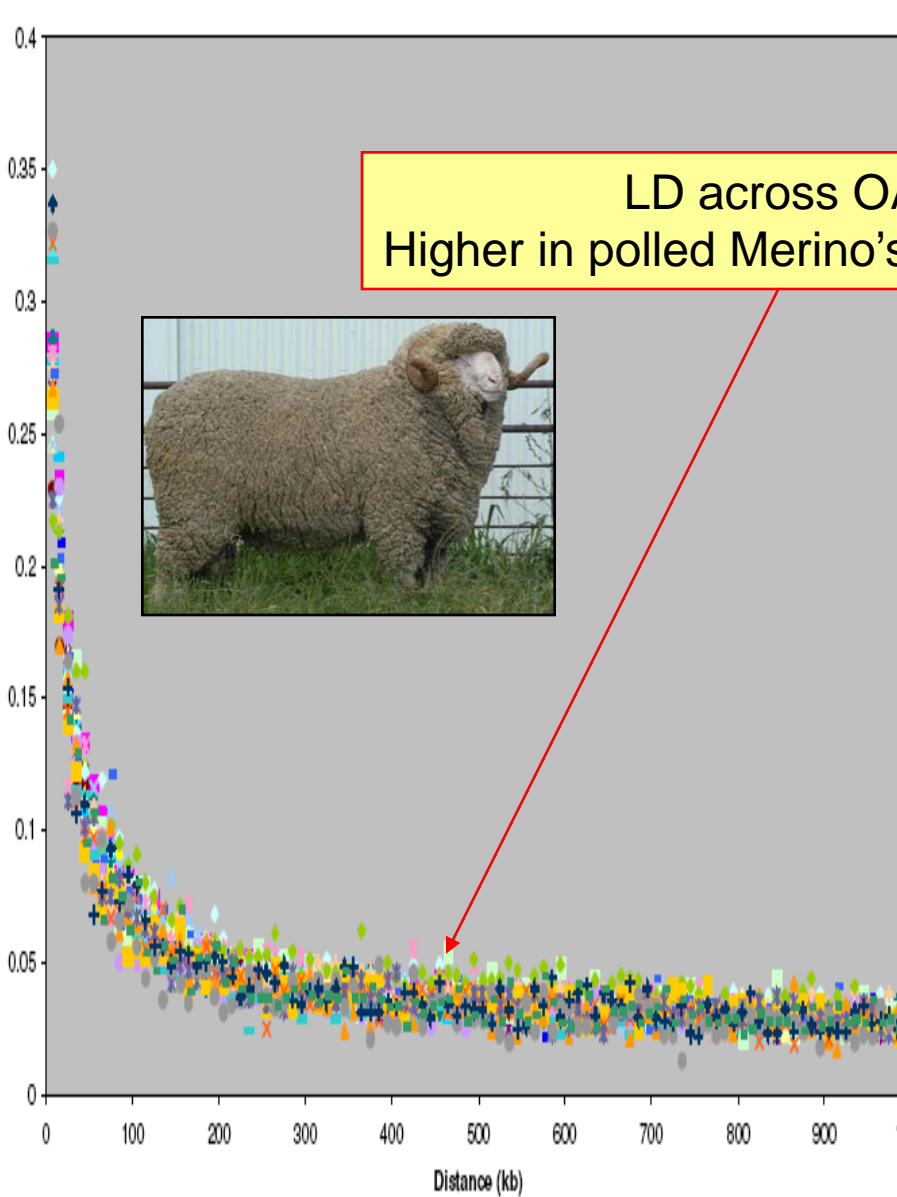
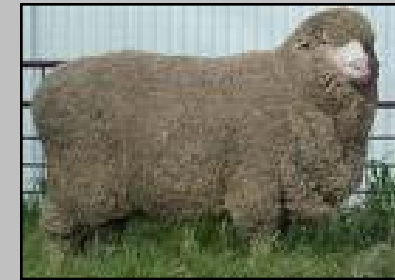


LD across chromosomes

Australian Merino

Australian Polled Merino

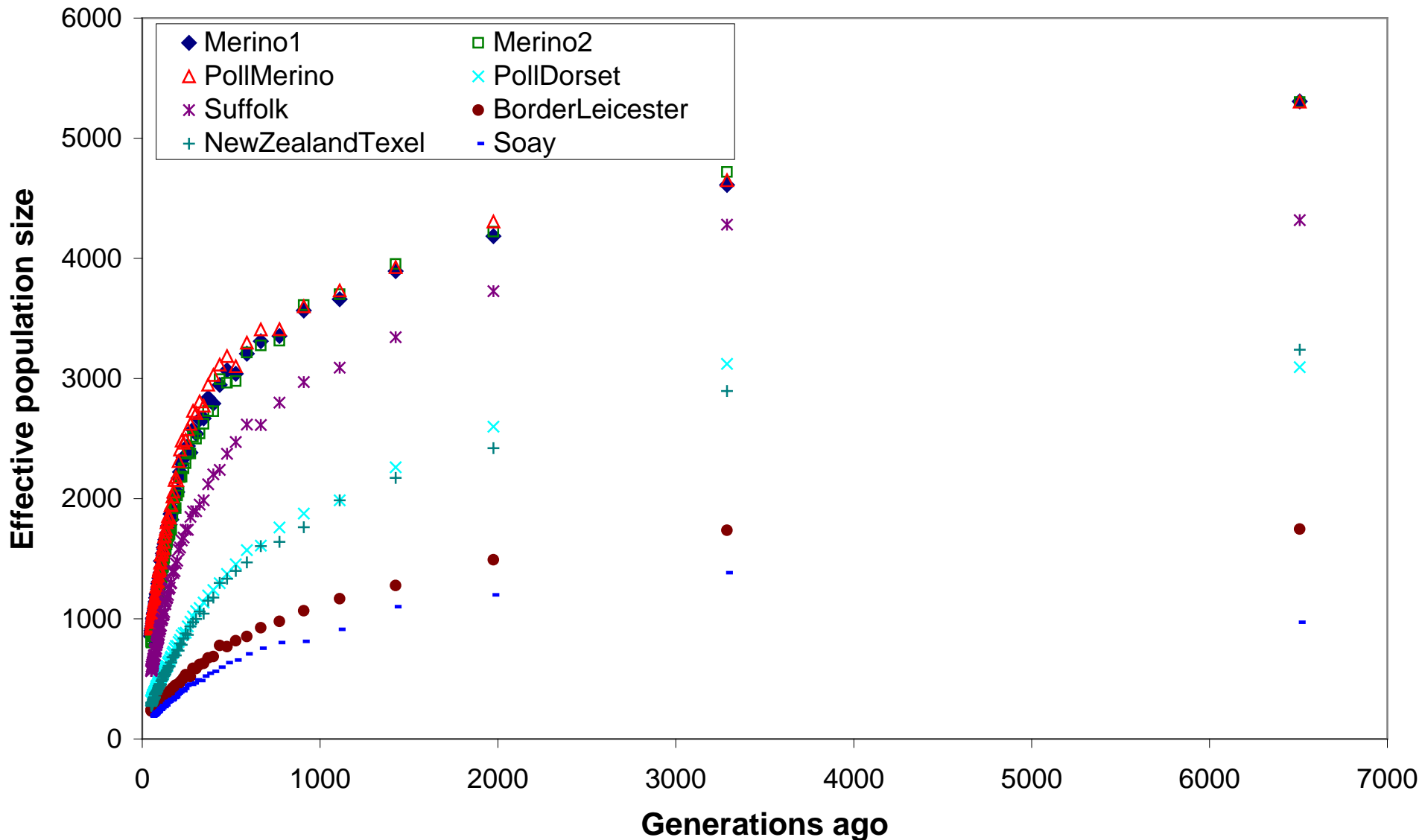
LD across OAR 10 (polled locus)
Higher in polled Merino's vs non OAR10 and non polled



- oar1
- oar2
- oar3
- oar4
- oar5
- oar6
- + oar7
- oar8
- oar9
- oar10
- oar11
- oar12
- oar13
- oar14
- oar15
- + oar16
- oar17
- oar18
- oar19
- oar20
- oar21
- oar22
- oar23
- oar24
- + oar25
- oar26

Population characteristics

Effective population size N_e



Method by Hayes et al (2003) Genome Research

Effective population size ~ 50 generations ago

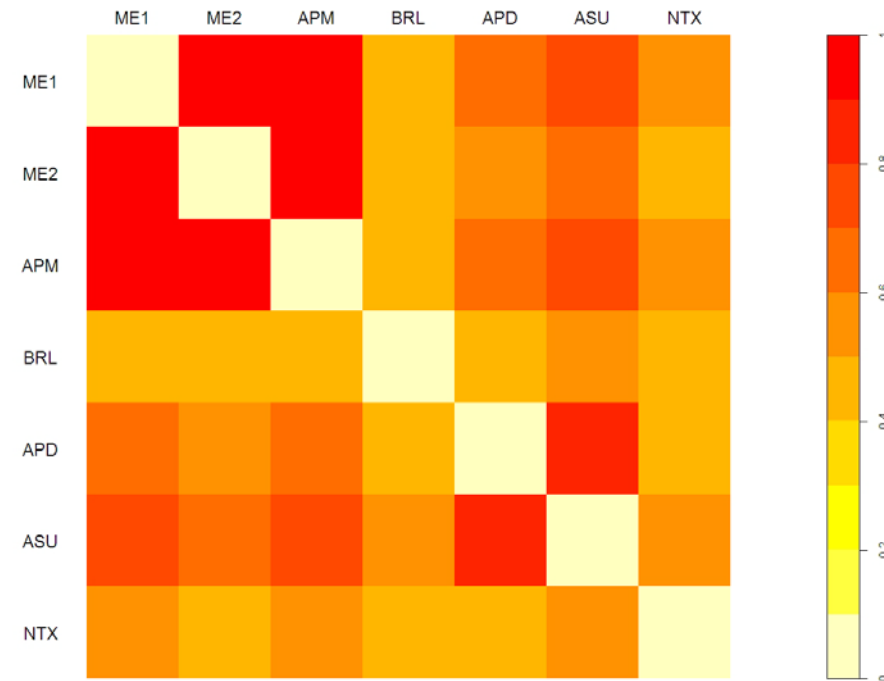
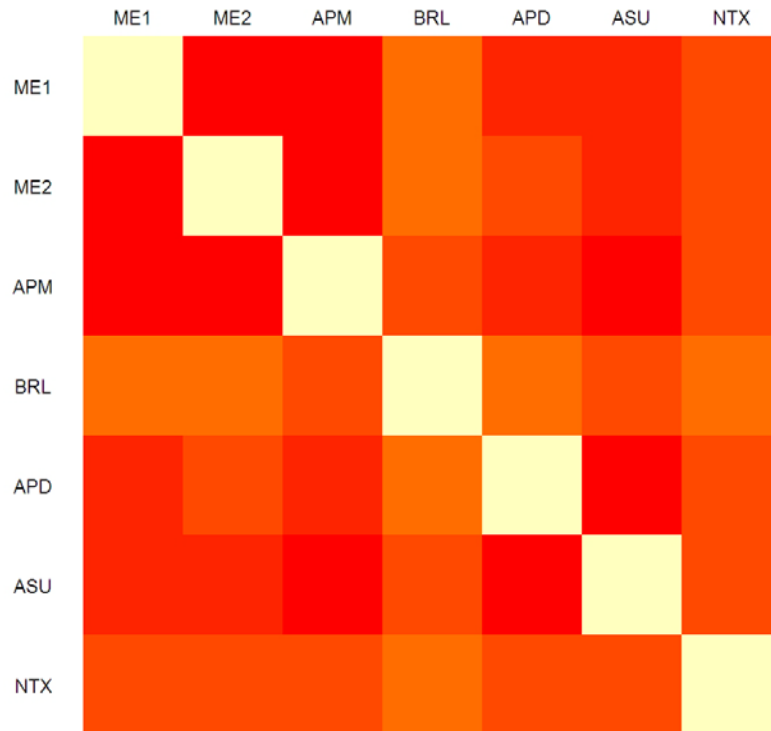
AfricanDorper	264	Comisana	589	NewZealandRomney	405
Afshari	465	CyprusFatTail	203	NewZealandTexel	282
Altamura	469	Deccari	837	Norduz	285
AustralianIndustryMe	853	DorsetHorn	134	Ojalada	588
AustralianMerino	833	EastFriesianBrown	186	Qezel	1317
AustralianPollDorset	318	EngadineRedSheep	482	Rambouillet	709
AustralianPollMerino	918	EthiopianMenz	1260	Rasaaragonesa	780
AustralianSuffolk	569	Finnsheep	795	RedMaasai	1070
BangladeshiBGE	582	Galway	322	Sakiz	165
BangladeshiGarole	437	Garut	337	Santalnes	520
BarbadosBlackBelly	374	GermanTexel	448	SardinianAncestralBI	306
BlackHeadedMountain	247	GulfCoastNative	611	ScottishBlackface	528
BorderLeicester	242	IndianGarole	364	ScottishTexel	305
BrazilianCreole	371	IrishSuffolk	300	Soay	194
BundnerOberlanderShe	253	Leccese	556	Spael-white	339
Castellana	546	MeatLacaune	835	Sumatra	336
Changthangi	684	Merinolandschaf	407	SwissBlack-BrownMoun	333
ChineseMerino	397	MilkLacaune	714	SwissMirrorSheep	262
Chios	334	Moghani	825	SwissWhiteAlpineShee	374
Churra	600	MoradaNova	327	Tibetan	903
				ValaisBlacknoseSheep	288
				ValaisRedSheep	173
				Wiltshire	100

Population divergence

LD between breeds @10kb

LD between breeds @50kb

B

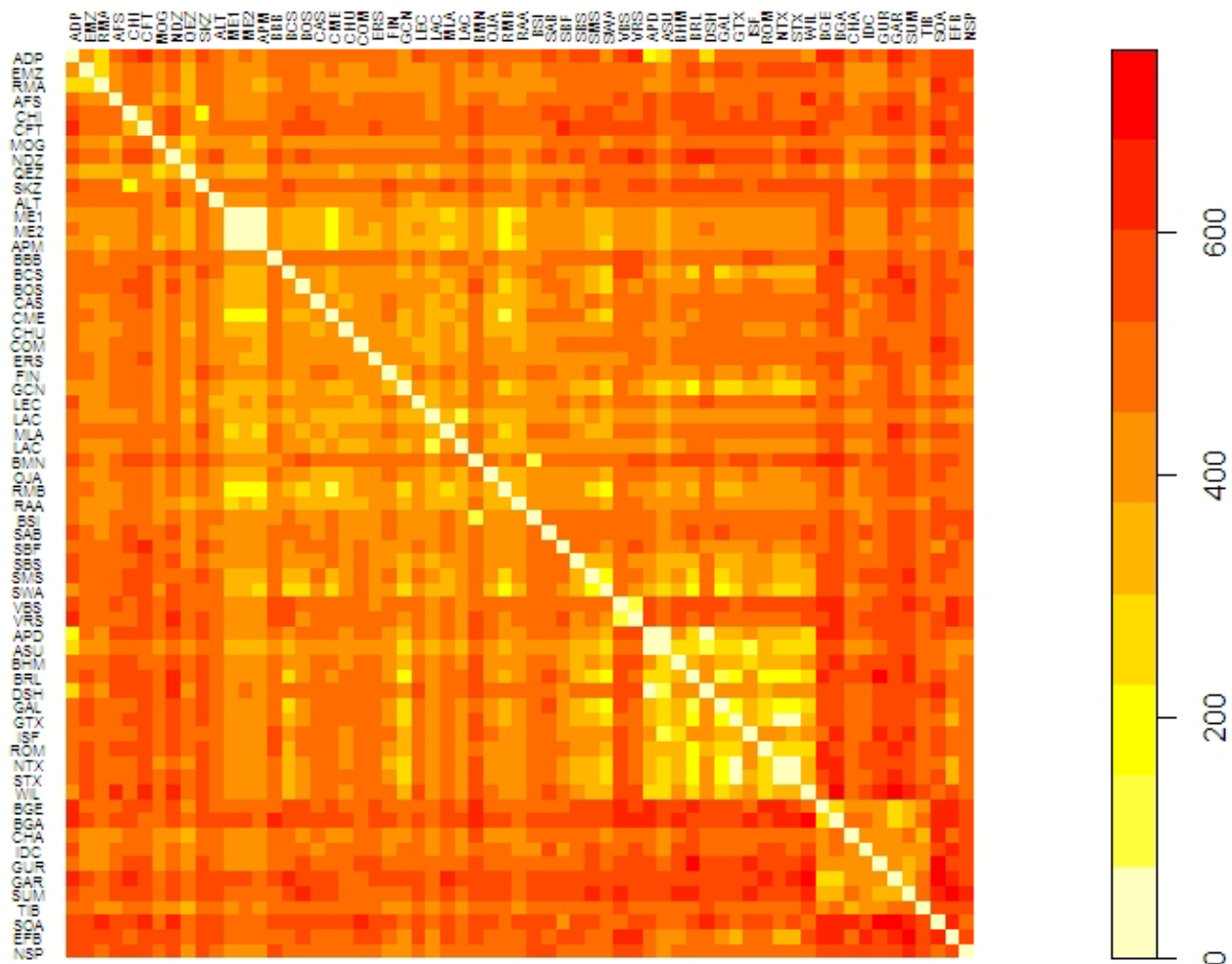


Mer 1 = Merino 1
Mer 2 = Merino 2
APM = Aus Poll Merino

BRL= Border Leicester
APD= Aus Poll Dorset
ASU= Aus Suffolk
NTX= NZ Texel

Divergence times estimated from LD sharing

Divergence time in generations



Divergence between human populations estimated from linkage disequilibrium.

Sved JA, McRae AF, Visscher PM. *Am J Hum Genet.* 2008 Dec;83(6):737-43

Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle.

de Roos AP, Hayes BJ, Spelman RJ, Goddard ME. *Genetics.* 2008 Jul;179(3):1503-12. Epub 2008

Haplotype formation

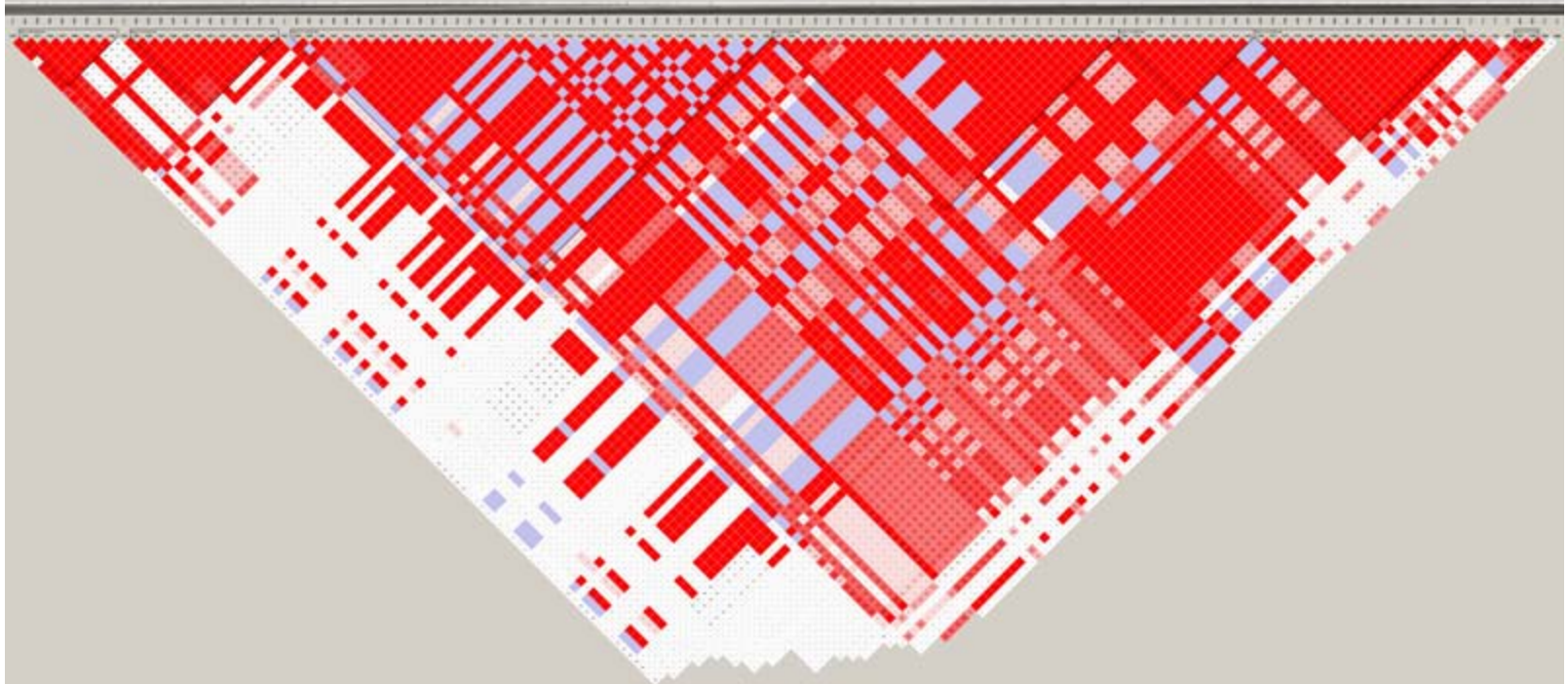
Used Haploview (Barrett et al 2005)

D` all pairwise combinations of SNP within chromosomes

Strong LD if 95% CI= 0.7 – 0.98

Define haplotype blocks based on SNP in strong LD

Soay Chr 2 section



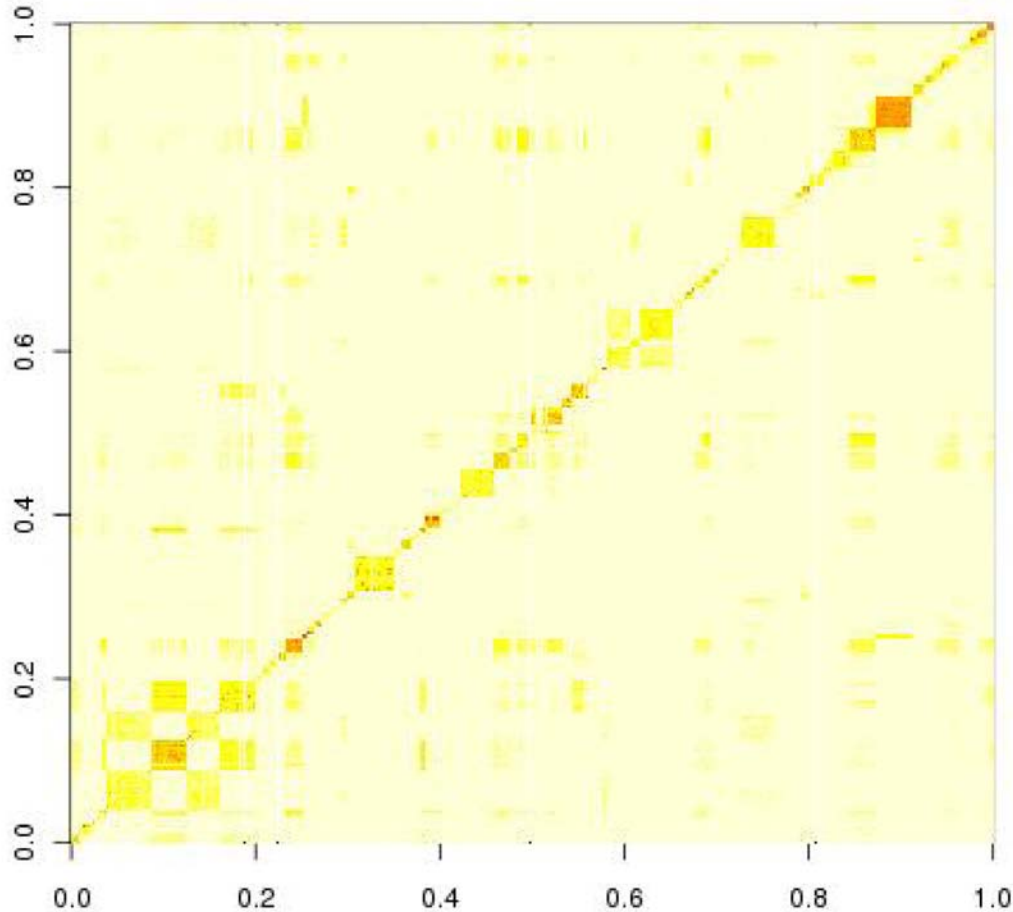
Churra Chr 2 section



Haplotype coverage genome

breed	breed_code	n	n block 3+ SNPs	Mean kb (3+SNP)	Min kb (3+SNP)	Max kb (3+SNP)	% cov (3+SNP)
Churra	CHU	120	169	126.1	14.1	986.4	0.8
Soay	SOA	110	1619	357	12.7	3547.9	21.84
Australian Suffolk	ASU	109	198	138.1	11.2	1623.1	1.03
Australian PollDorset	APD	108	439	177.3	11.2	1923.6	2.94
MilkLacaune	LAC	103	168	141.9	12	834.6	0.9
Rambouillet	RMB	102	190	138.8	11.2	858.1	1

Fig. Relationship between animals based on haplotype similarity.

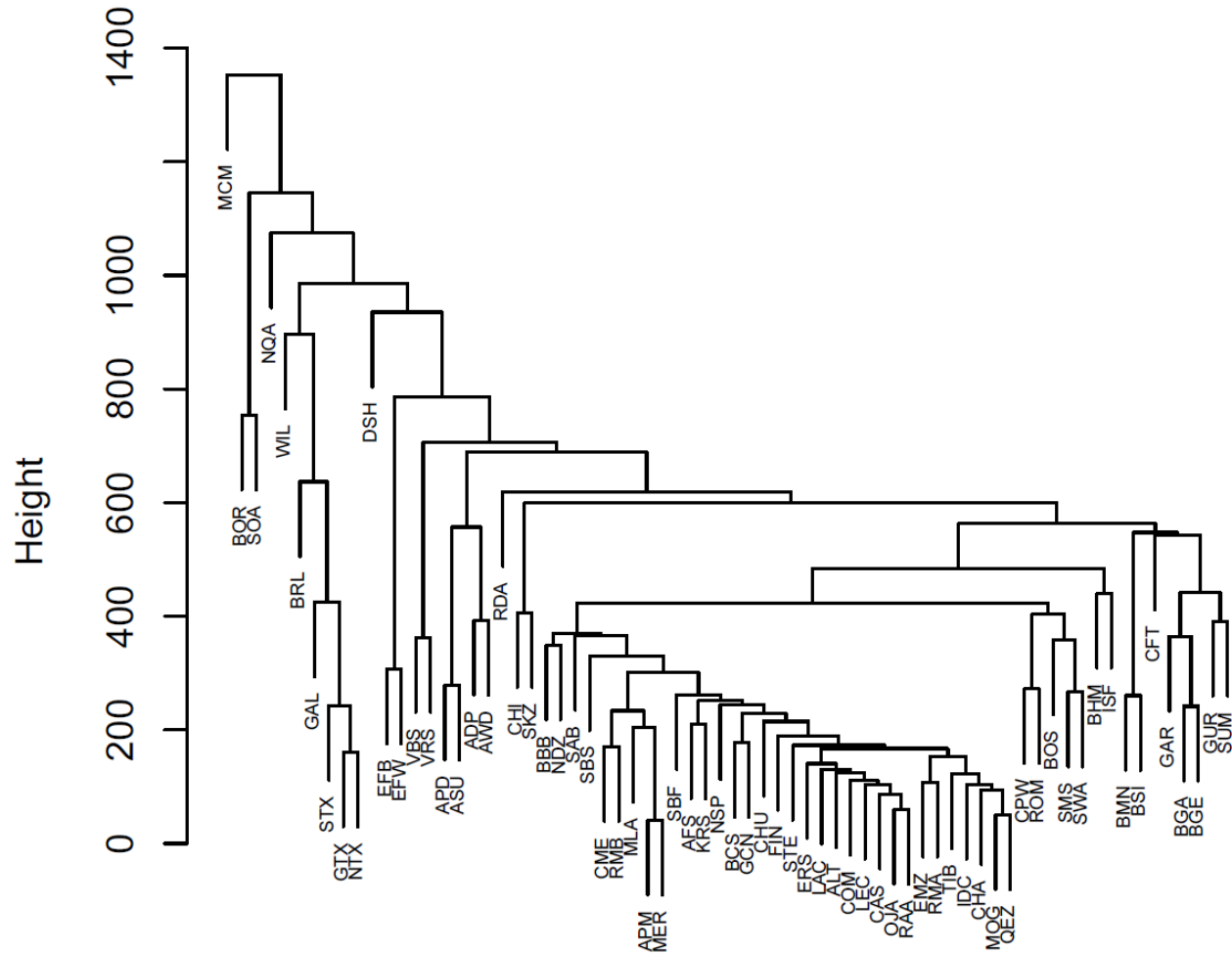


Haplotype similarity

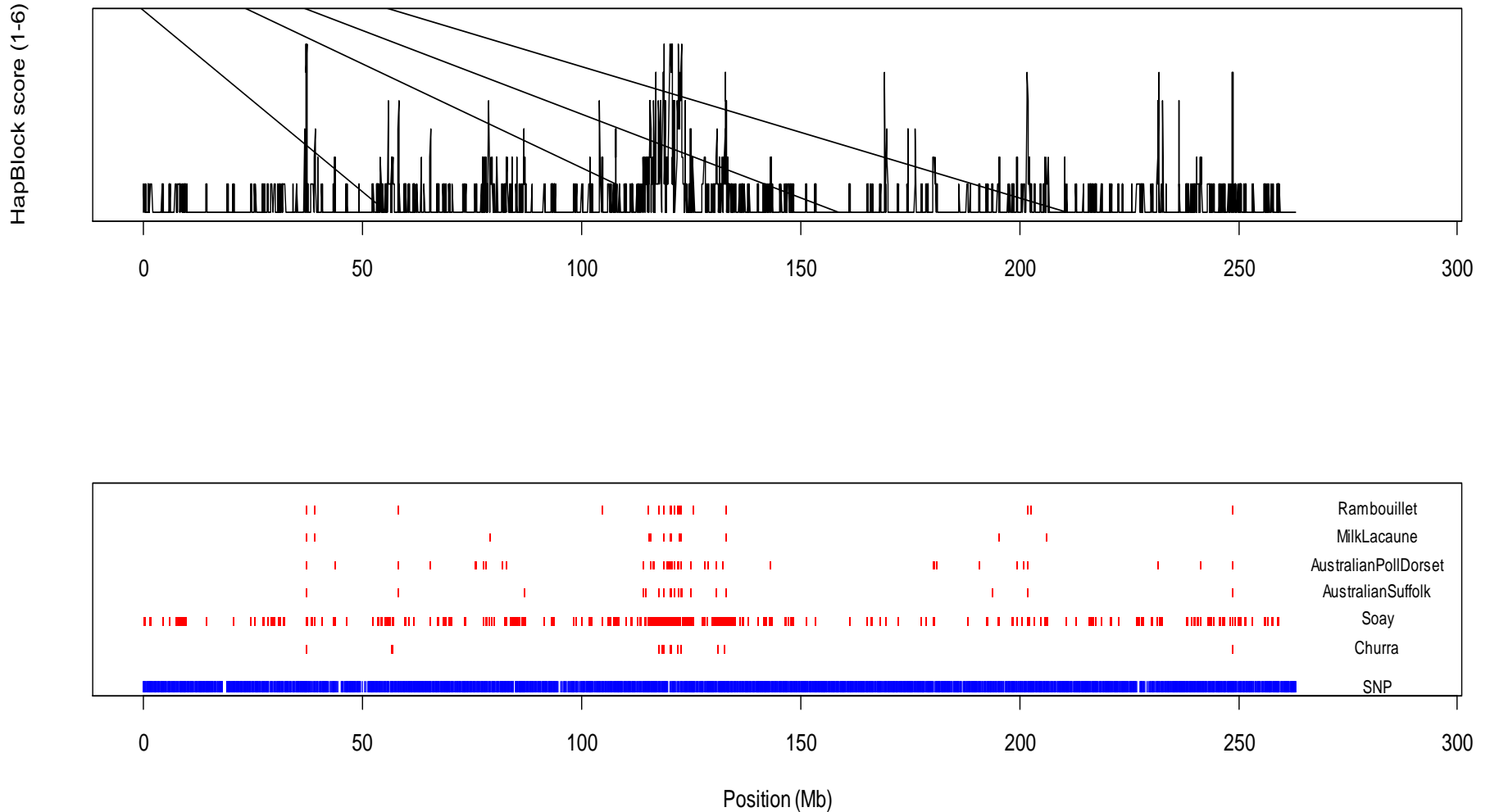
- IBD shared segments
- (GERMLINE- Gusev etal 2009)
- Length and number of segments
- each pair of individuals
- 2819 x 2819 matrix
- relatedness within breeds and between breeds

Breed clustering based on haplotype similarity

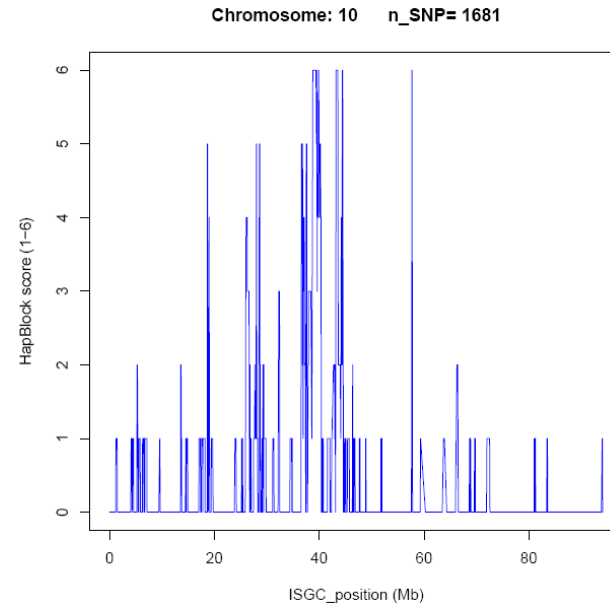
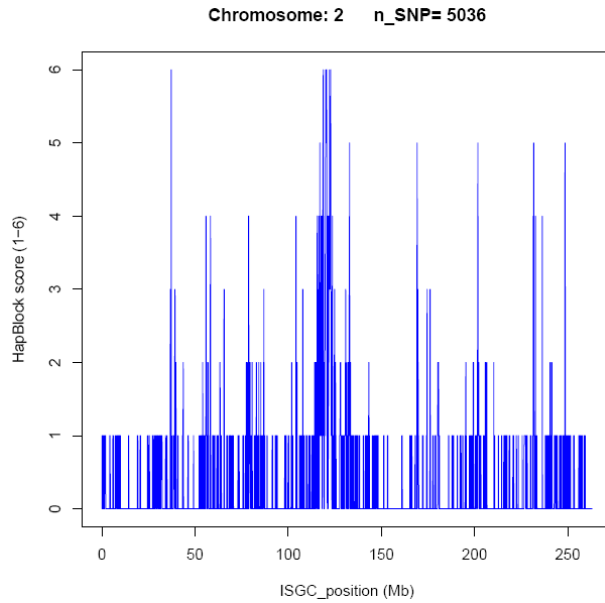
Cluster Dendrogram



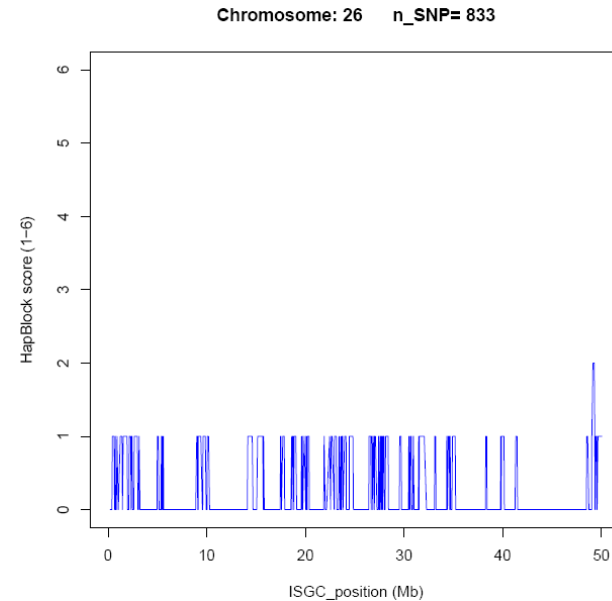
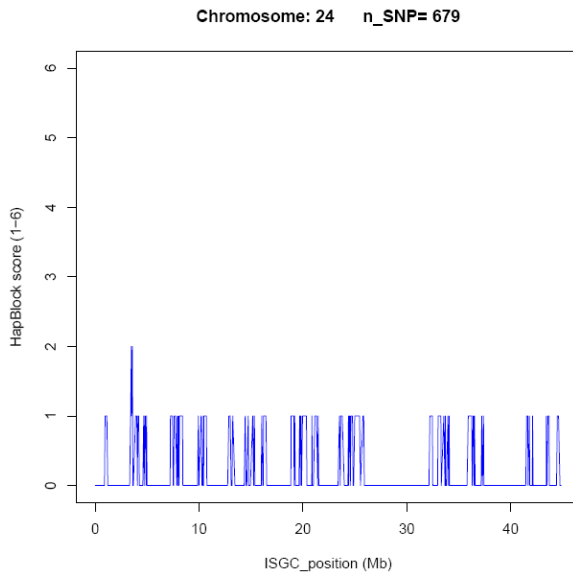
HapMap of six breeds: chromosome 2 occurrence of haplotype blocks across genome



Occurrence of haplotype blocks across 6 breeds



Regions of high and low haplotype formation



LODE- mapping markers by LD

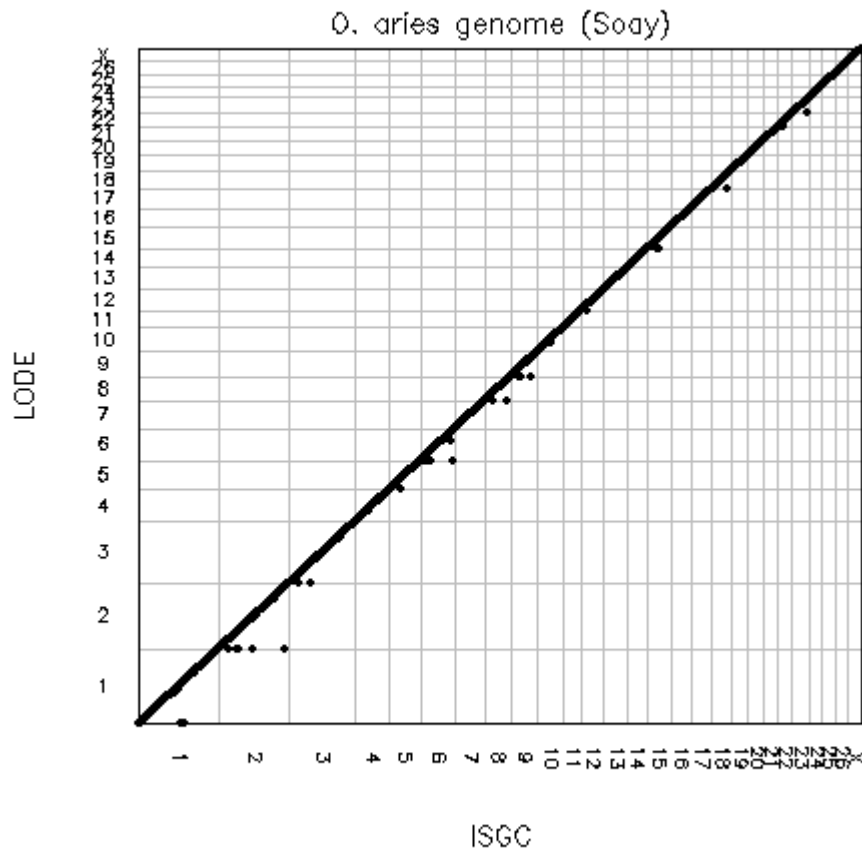
Khatkar et al 2010

Locus Ordering by Dis-Equilibrium

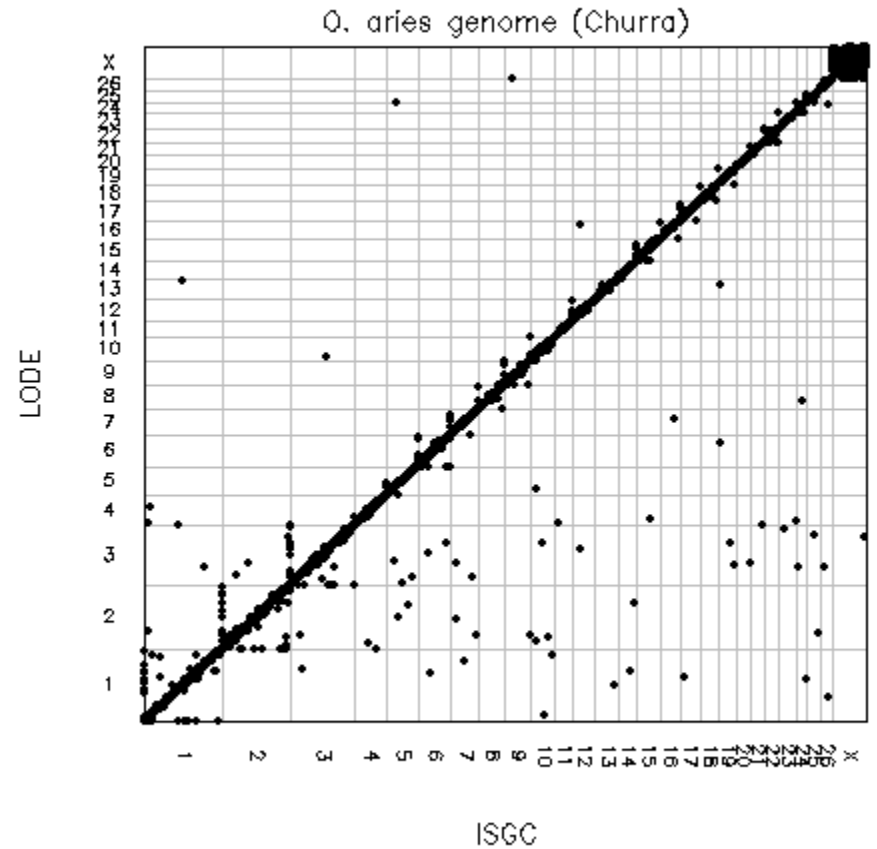
- placing markers by LD in relation to others
- only SNP MAF>0.01 considered
- r^2 calculated each SNP to all other SNP (mapped and **unmapped**)
- SNP assigned to chromosome – max r^2 and number hits $r^2>0.1$
- SNP assigned to position- adjacent to mapped SNP with highest r^2

- Each SNP taken as “unmapped” in turn 49,000 x 49,000
- positioned based on LD in two breeds – Soay and Churra
- Mapped position compared against ISGC Virtual Sheep Genome (VSG)
- VSG contains integrated map all known markers & comparative genome

Agreement marker positioning LOD vs Virtual Sheep genome



Success in breed with HIGH LD



Success in breed with LOW LD

Concluding observations

- LD profiles in sheep breeds similar
- LD relatively low in sheep based on ovine 50K beadchip
- LD sufficient to detect divergence time
- LD short and long range sufficient to map breed development
- LD sufficient to pick up evidence of selective sweeps
- LD sufficient to validate marker maps/possible assembly errors

- Limited genome coverage in haplotype blocks
- N_e in sheep appears quite high most breeds

- Q do we need higher SNP density array??????????????

Current development sheep genome

- reference genome sequence
 - 30x Roslin Institute- Texel ram – BAC library
 - 15X paired end 75 bp read 250-300bpfragment- in progress
 - 10X mate pair 1.5kp insert - planned
 - 1X mate pair 8kb insert 454 150bp ends reads (Baylor)
 - 1x mate pair 20kb insert 150 bp end reads 454(Baylor)
- Assembly- CSIRO- Brisbane + Agresearch NZ+.....
- Funding – ISGC (ISL) + USDA- Noelle Cockett Utah State U.
 - AgResearch
 - Baylor College of Medicine
 - Genesis Farraday

Acknowledgements

John McEwan
Rudi Brauning
Theresa Wilson
Alan McCulloch



Marylinn Munson
Kim Gietzen
Christian Haudenschild



James Kijas
Brian Dalrymple
Wes Barris
David Townley
Sean McWilliam
Russell McCulloch



Peter Wilson
Annette McGrath



Richard Gibbs
Donna Muzny

Jo-Ann Stanton



An Australian Government Initiative
Backing Australia's Ability



United States Department of Agriculture



The University of Sydney



Jill Maddox (U. Melb)
Herman Raadsma (U. Syd)
Frank Nicholas (U. Syd)
Hutton Oddy (UNE)
Chris Warkup (GF)
Noelle Cockett (USU)

Tim Smith (USDA)
Mike Heaton (USDA)
Curt van Tassell (USDA)
Terry Longhurst (MLA)
Mohammad Shariflou (U. Syd)
Ben Hayes (Vic DPI)

And finally.....

International Sheep Genomics Consortium

The screenshot shows a Windows Internet Explorer browser window displaying the International Sheep Genomics Consortium website. The browser's address bar shows the URL <http://www.sheephapmap.org/>. The website's header features the ISGC logo, which includes the text "ISGC" and "International Sheep Genomics Consortium" alongside a stylized illustration of a sheep and a lamb. A search bar labeled "ISGC Search" is positioned to the right of the logo. Below the header is a navigation menu with the following items: Home, About, Project Details, Partner Resources, and Contact. A large black text box is overlaid on the page, containing the following text in blue font: "Better, faster cheaper than Cattle Chicken and Pigs", "Learn from others", "Do what is best", and three URLs: www.sheephapmap.org, www.livestockgenomics.csiro.au/sheep/oar1.0.php, and isgcdata.agresearch.co.nz. The browser's status bar at the bottom shows "Done" and "Internet" with a 100% zoom level.