Realized autozygosity and genetic differentiation using Landrace x Large White crossbreds

L. Gomez-Raya[1,2], WM Rauw[1,2], J Dunkelberger[2],
and JCM Dekkers[2]

gomez.luis@inia.es

Instituto Nacional de Investigaciones
Agroalimentarias[1], Madrid, Spain.

Department of Animal Science[2]
Iowa State University
USA

Genetic Analyses of Crossbreds

- Swine breeds genetically separated for many generations are routinely crossed to take advantage of heterosis

- Analysis of crossbred genotypic information can provide information on common (autozygosity) and divergent chromosomal regions (F_{ST}) in the genomes of the parental breeds

OBJECTIVES

1) determining levels and regions of autozygosity as highlighted by Runs of Homozygosity (ROH).

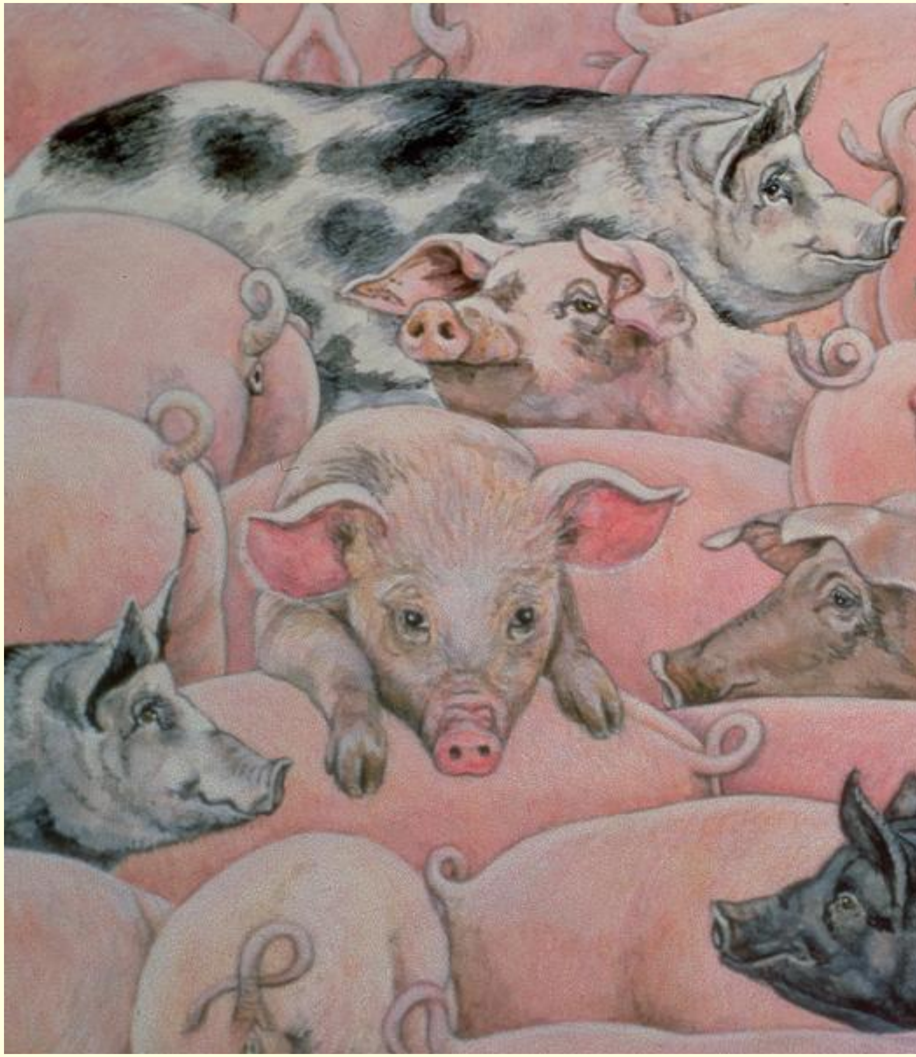
2) determining the genomic distribution of Δ (difference in allele frequency in the parental breeds) and F_{ST} of Wright in the two parental breeds.

A method to estimate allele frequencies in the two parental breeds when only information in the crossbred was developed.

ANIMAL MATERIAL

- 1173 barrows from a Landrace x Large White Crossbreds from six field trials and two companies.
- Illumina porcine SNP array with 50K or 60 K SNP (after editing 32,659 SNPs polymorphic common to all individuals).

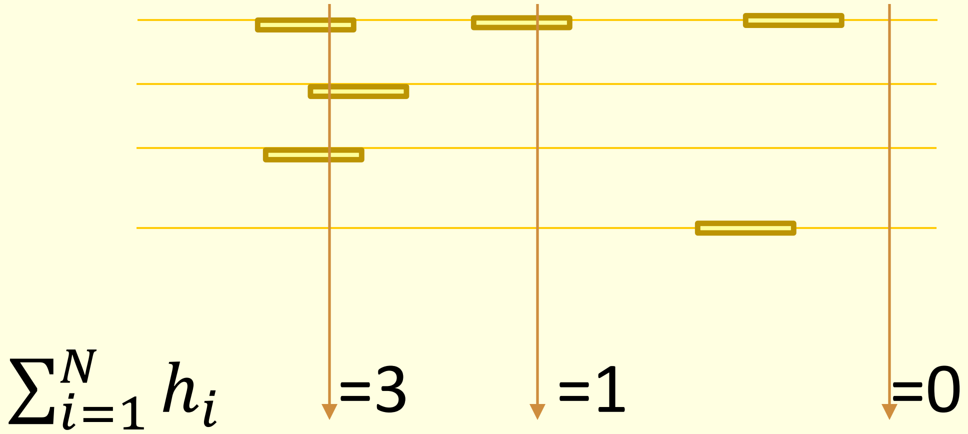
ROHs (25 homozygote SNPs) was determined for all ind.



Autozygosity

Probability for a SNP of being autozygous (Kim et al)

$$F_L = \frac{\sum_{i=1}^{N} h_i}{N}$$



where h_i is the ROH appraisal of the *i-th* individual (h_i =1 if the *i-th* individual contains a ROH at that position and h_i = 0, otherwise).

Max Likelihood to test autozygosity

The likelihood of the number of animals with ROH status h_i =1 can be thought as following a binomial density with probability F_L:

$$L(F_L) = \frac{N!}{m!\,(N-m)!}(F_L)^m(1-F_L)^{N-m}$$

where m is the number animals with status h_i =1 in a given SNP position. N=number of individuals.

Likelihood Ratio Test

$$
\begin{aligned}
LRT = 2[&m\, ln(\hat{F}_L) \quad + (N-m)\ln(1-\hat{F}_L) \\
&-m\, lnF_0^{ROH} - (N-m)\, ln(1-F_0^{ROH})]
\end{aligned}
$$

where $F_0^{ROH} = \frac{\sum_{i=1}^{N} F_i^{ROH}}{N\, n_{snp}}$

Null hypothesis

Equal Prob. of autozygosity for any position in the chromosome

Determining the genomic divergence in parental breeds

F_{ST} method of Wright for each SNP :

$$F_{ST} = \frac{\sigma_f^2}{f(1-f)}$$

where *f* is the average of $\widehat{f_A}$ and $\widehat{f_B}$ and σ_f^2 its variance.

Difference in allele frequency in the parental breeds)

Δ = abs($\widehat{f_A} - \widehat{f_B}$)

Estimation of allele frequencies in parental breeds using crossbred information

$$L(f_A, f_B | N_{11}, N_{12}, N_{22}) = K\,(f_A f_B)^{N_{11}} \left(f_A(1 - f_B) + f_B(1 - f_A)\right)^{N_{12}}$$

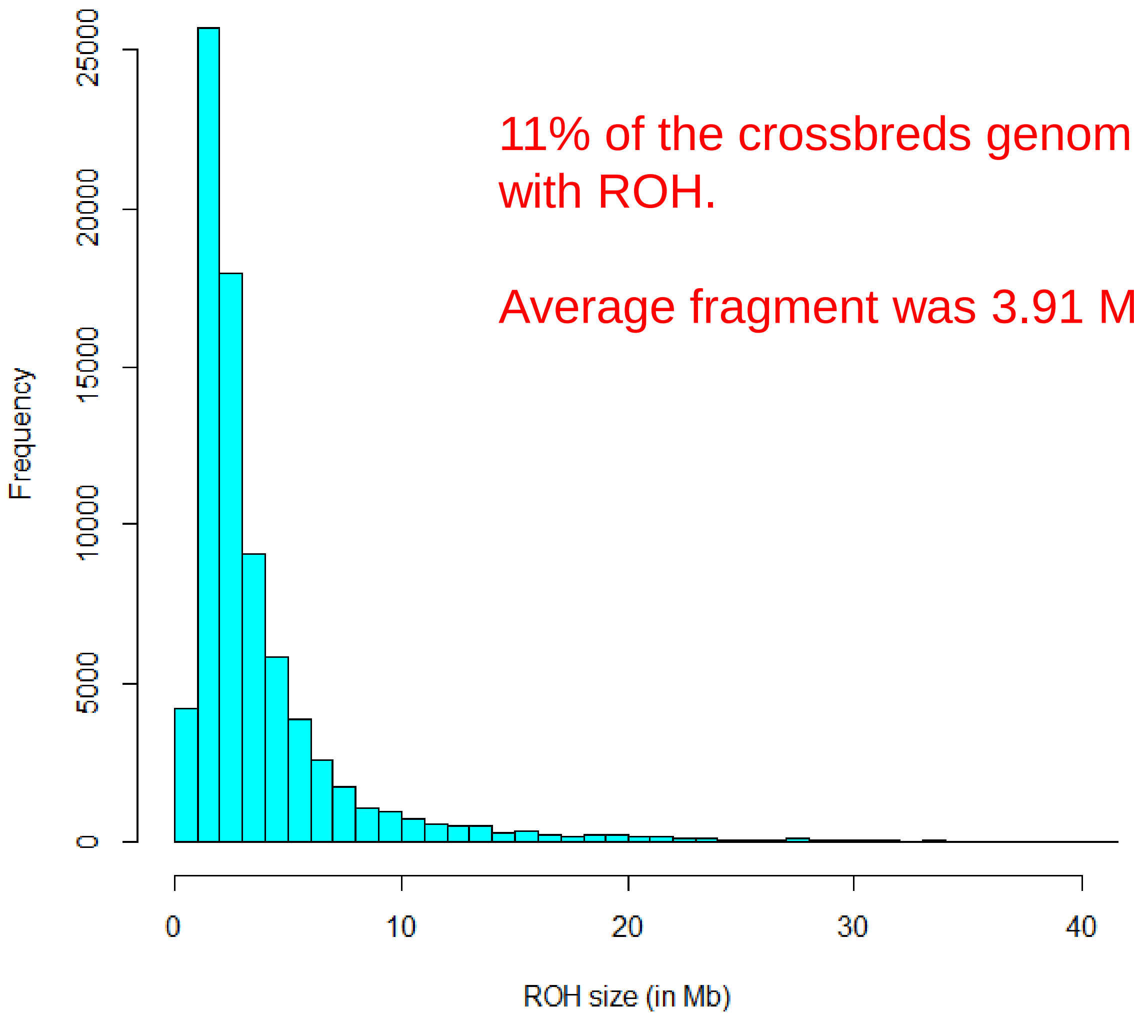
$$((1 - f_A)(1 - f_B))^{N_{22}}$$

$$LRT = 2[\ln L\,(f_A^{Max}, f_B^{Max} | N_{11}, N_{12}, N_{22}) \quad - \ln L\,(f_A^{o}, f_B^{o} | N_{11}, N_{12}, N_{22})]$$

f_A^{Max} and f_B^{Max} are the allele frequencies estimates in breeds A and B

RESULTS

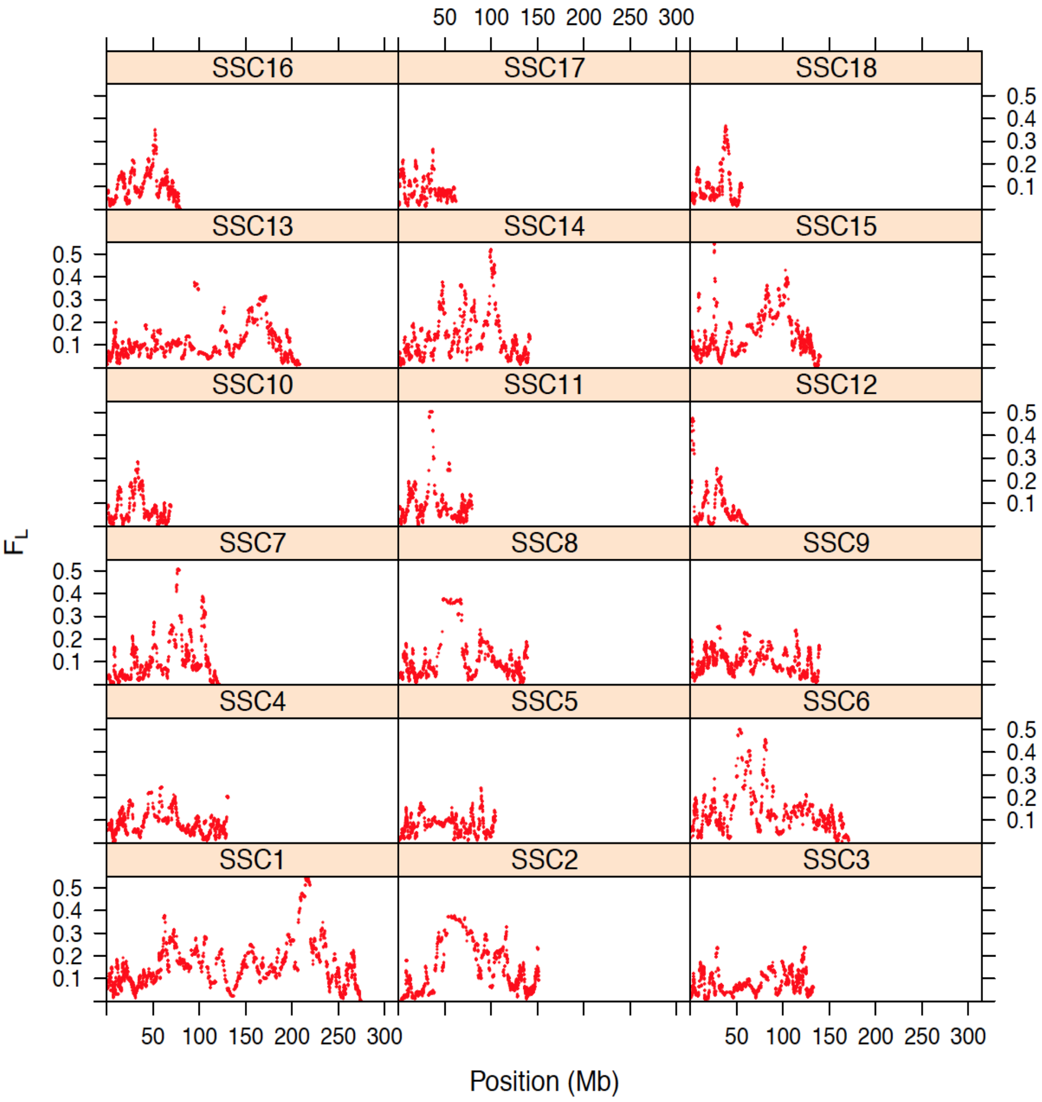
Distribution of the size of ROH among 1173 Landrace x Large-White x crossbreds



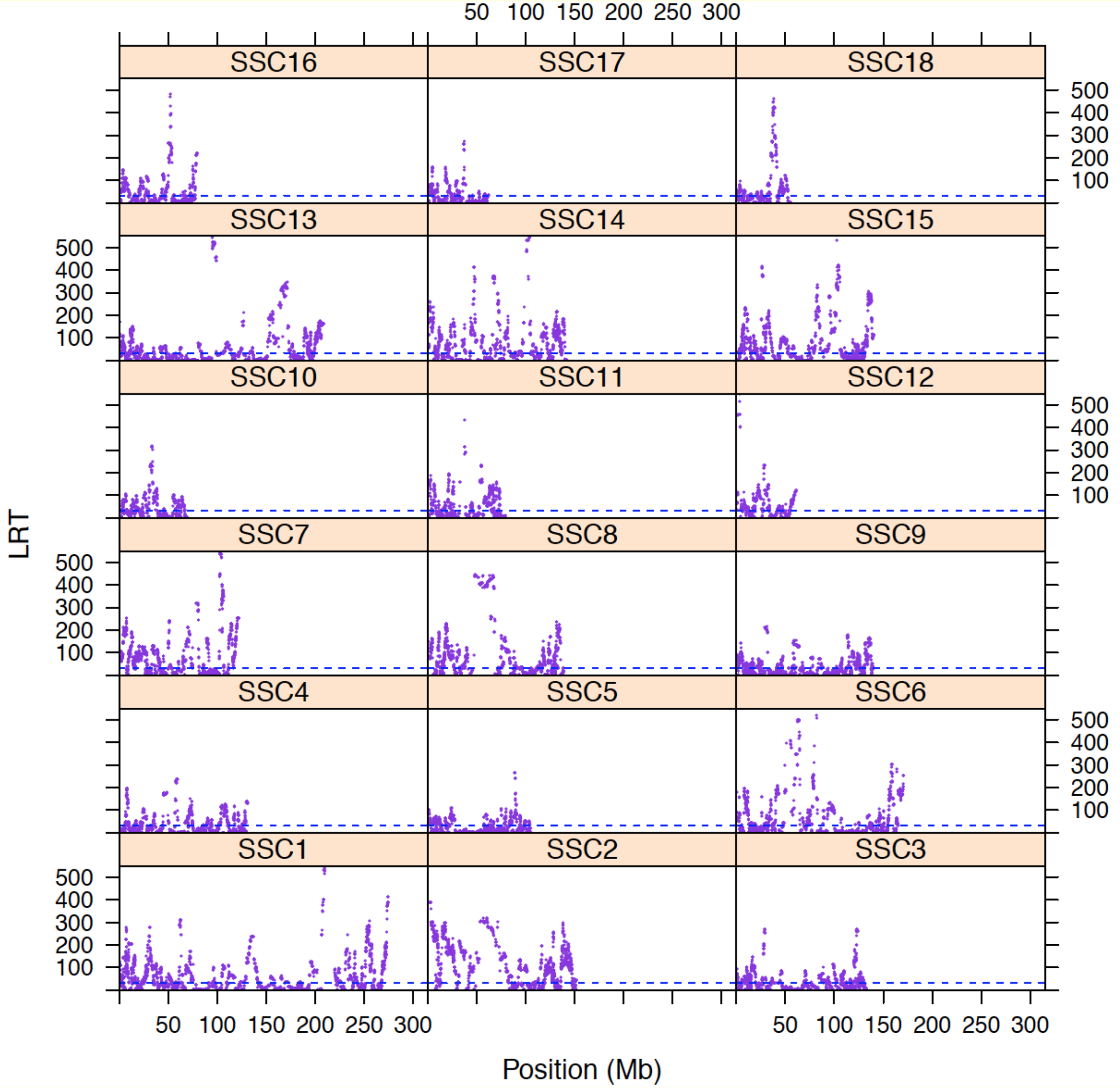
11% of the crossbreds genome is covered with ROH.

Average fragment was 3.91 Mb.

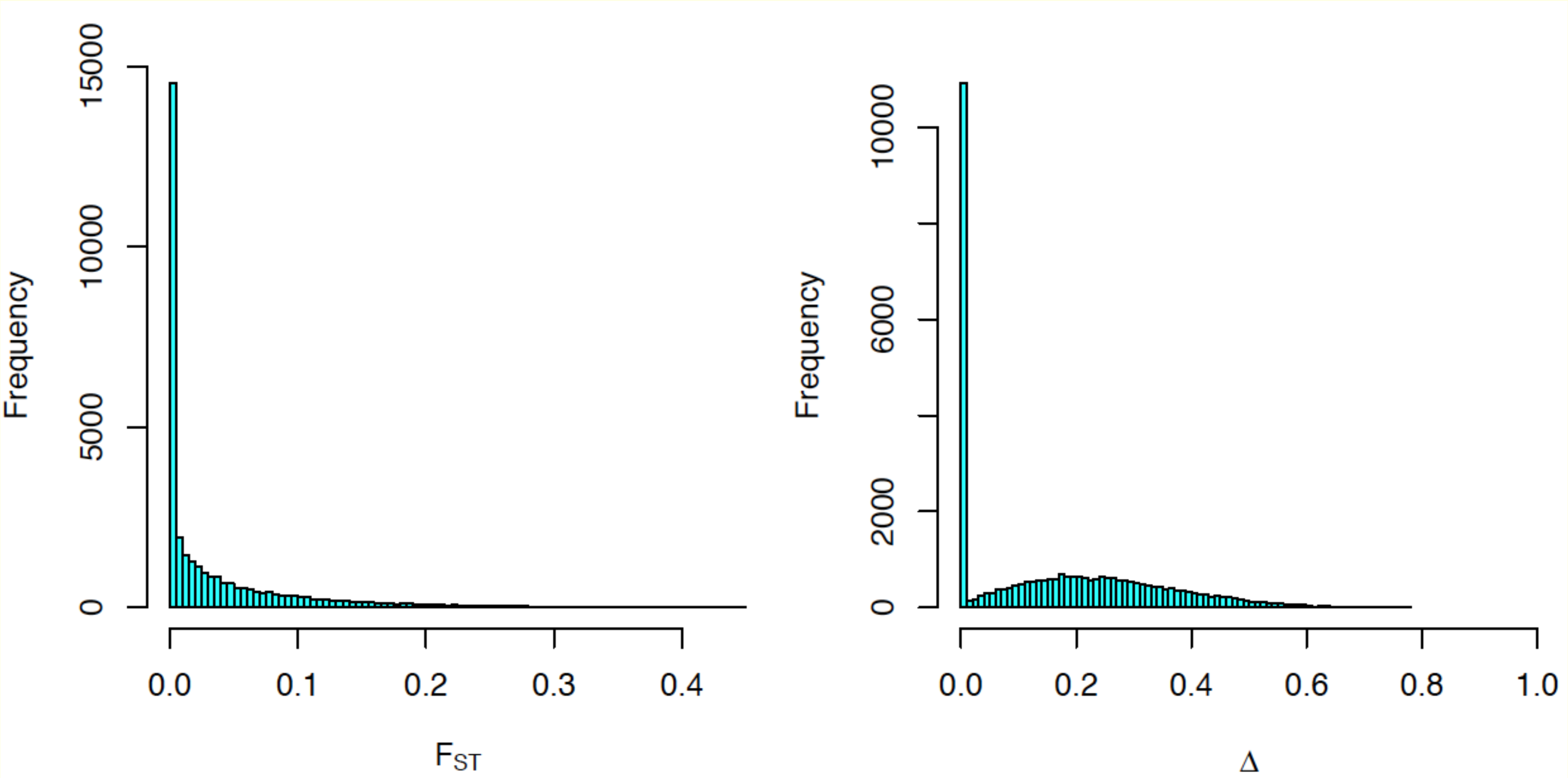
Map of autozygosity of Landrace x Large White crossbreds



Manhatan plot of F_L



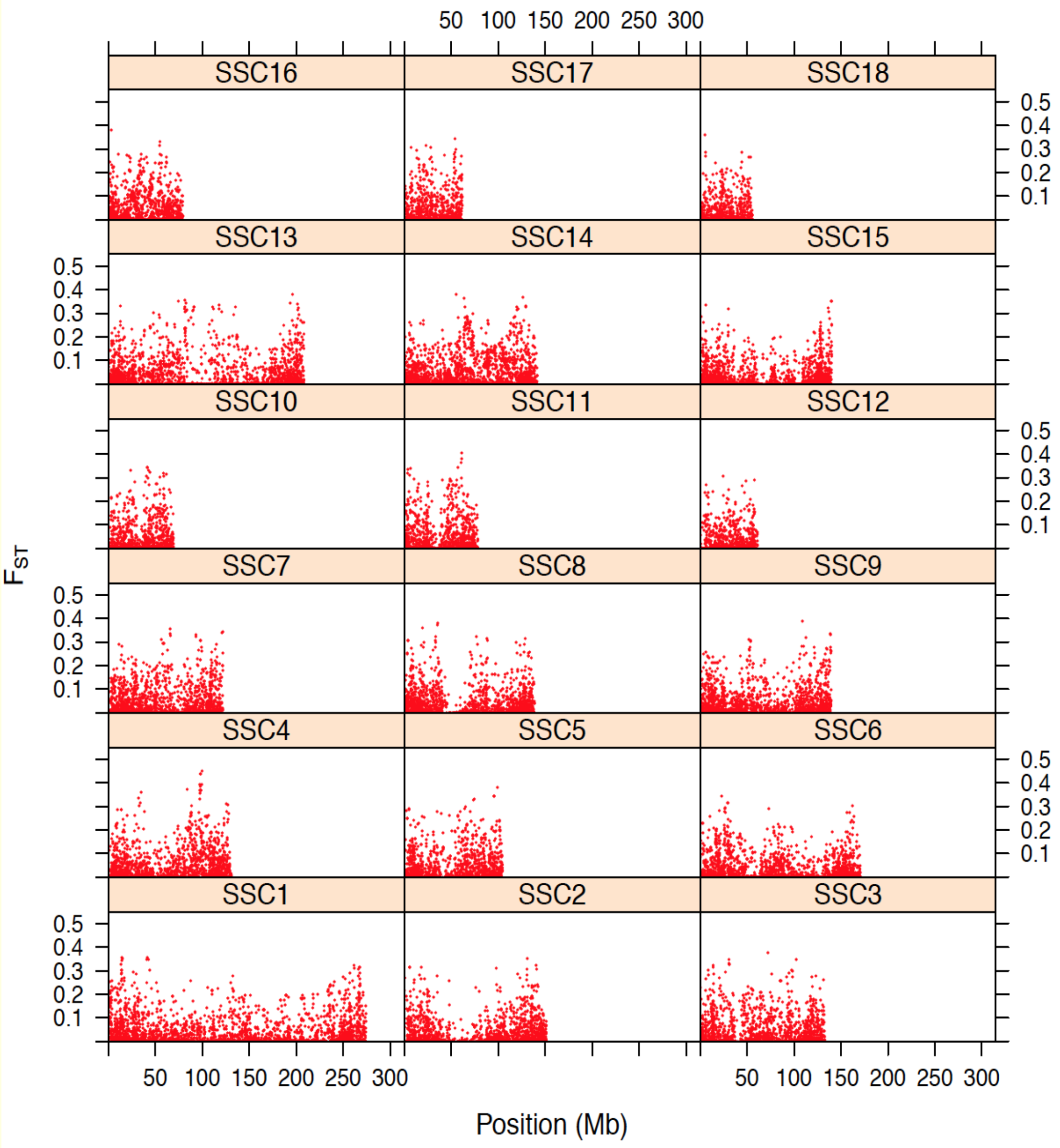
Distribution of F_{ST} and Δ across the genome of Large-White x Landrace cross



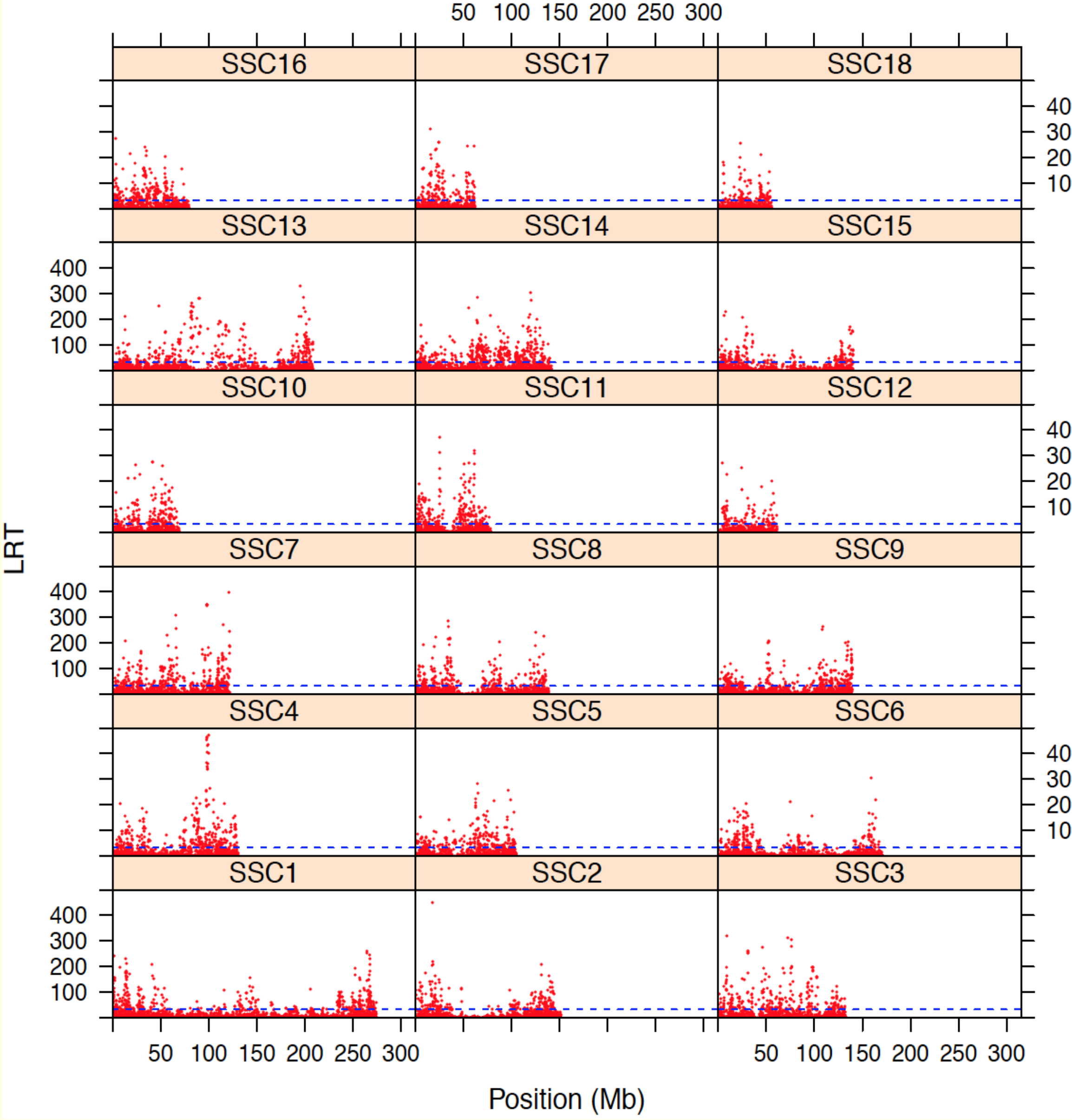
About 50% of the loci are segregating at the same frequency in the two breeds

Average F_{ST} across the genome was 0.038 (SD=0.059)

Map of divergence of Landrace and Large White using the F_{ST} statistic.



Manhattan plot of Δ of Landrace x Large White crossbreds



Correlation and parameter estimates between F_L of SNP distance to the closest telomere

Model

- I (constant relationship)
- II (proportional to the telomere distal distance)
- III (proportional to the value of F_L).

Chromosome	I	II	III
1	0.29	0.16	0.34
2	0.68	0.70	0.61
3	0.10	0.16	0.03
4	0.39	0.45	0.36
5	0.16	0.15	0.21
6	0.42	0.39	0.41
7	0.37	0.30	0.53
8	0.43	0.47	0.36
9	0.25	0.22	0.33
10	0.54	0.63	0.49
11	0.29	0.33	0.25
12	0.03	0.14	0.20
13	0.31	0.23	0.37
14	0.55	0.49	0.61
15	0.44	0.39	0.46
16	0.54	0.47	0.55
17	0.02	0.06	0.03
18	0.26	0.19	0.32

Correlation between SNP distance to the closest telomere and F_{ST} or Δ.

Chromosome	# SNPs	F_{ST}	Δ
1	3,406	-0.11	-0.10
2	2,160	-0.13	-0.17
3	1,942	0.12	0.18
4	2,078	0.03	0.02
5	1,587	0.02	-0.02
6	2,279	-0.03	-0.03
7	1,940	0.04	0.00
8	1,989	-0.01	-0.02
9	2,177	-0.06	-0.04
10	1,063	0.04	0.04
11	1,265	0.05	0.01
12	835	0.05	0.03
13	2,565	0.07	0.02
14	2,248	0.24	0.28
15	1,987	-0.14	-0.14
16	1,262	0.23	0.24
17	989	-0.02	-0.01
18	887	0.07	0.14
Overall	32,659	-0.00(0.006)	-0.02(0.006)

CONCLUSSIONS

ABOUT 11% of the GENOME OF CROSSBREDS IS COVERED WITH ROH

Autozygosity is larger in the inner part of chromosomes
(maybe due to physical characteristics of chromosomes)

About 50% of the loci are segregating at the same frequency in the two breeds

Δ and F_{ST} are not larger in the inner or distal parts of chromosomes.

GENETIC ANALYSES OF CROSSBREDS MAY REVEAL INFORMATION OF PARENTAL BREEDS

THANKS!

AKNOWLEDGEMENTS

Generation of these data was supported by the USDA ARS and NIFA award 2012-38420-19286.

Financial support from the Instituto Nacional de Investigaciones Agrarias and from the Salvador de Madariaga program of the Ministry of Education of Spain.

Regions of autozygosity in crossbreds after selecting the 200 SNP with the highest *LRT*

Chrom.	# Mar	Map v11		Map v10.2					
		start	end	start	end	Size (Mb)	F_L	Genes/transcripts	Function
1	53	214032391	219236113	192522425	196655180	5.20	0.53	SNORA31,FSCB,KLHL28, snoU13, TOGARAM1, PRPF39, SNORD127	Structural, reproduction
6	27	51038832	55086387	55106785	60112015	4.05	0.49	NLRP11,NLRP11,NLRP4 NLRP4,NLRP13,NLRP8 NLRP8,NLRP5	Structural, Inflammatory response
7	25	74920040	78406442	70022773	73348971	3.49	0.48	EAPP, SPTSSA, EGLN3, NPAS3, U6, GPR33, HEATR5A, HECTD1, AP4S1, 7SK	Structural
11	18	33021830	37259574	31389557	35109328	4.24	0.49	PCDH17	Nervous system
12	31	1610942	3289920	353286	569474	1.68	0.46	TBCD, B3GNTL1, METRNL	Structural
14	29	98383612	100011521	90372482	91598259	1.63	0.5	5S_rRNA, SNORA31	Structural
15	17	25554685	25918412	22180635	22522196	0.36	0.53	ENSSSCG00000015707, ENSSSCG00000015708	Not Characterized Structural

Using the over representation test of software PANTHER they were significant associated to

1) antigen processing and presentation of peptide or polysaccharide antigen via MHC class II,

2) induction of apoptosis, and 3) cellular defense response.