

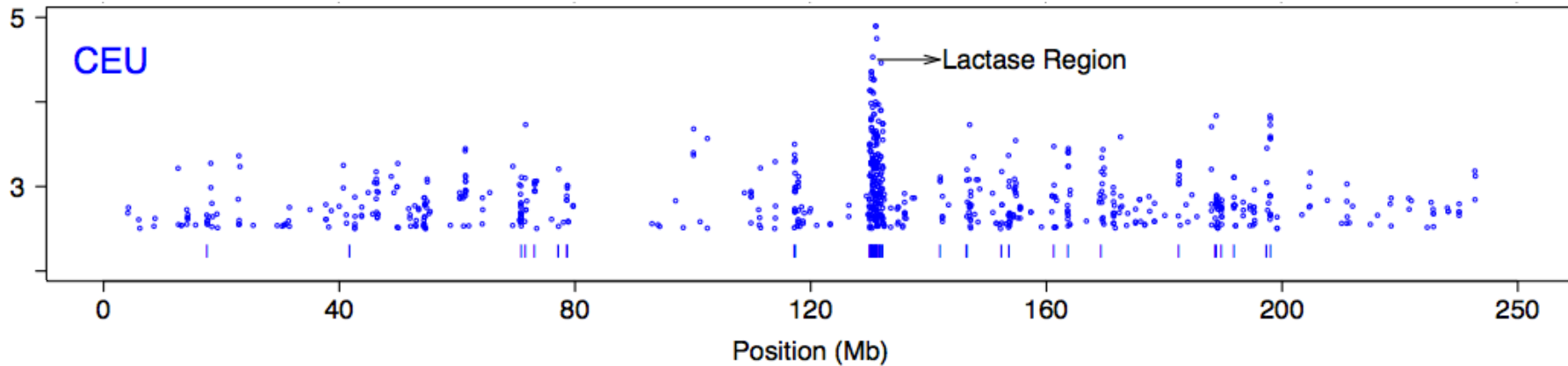
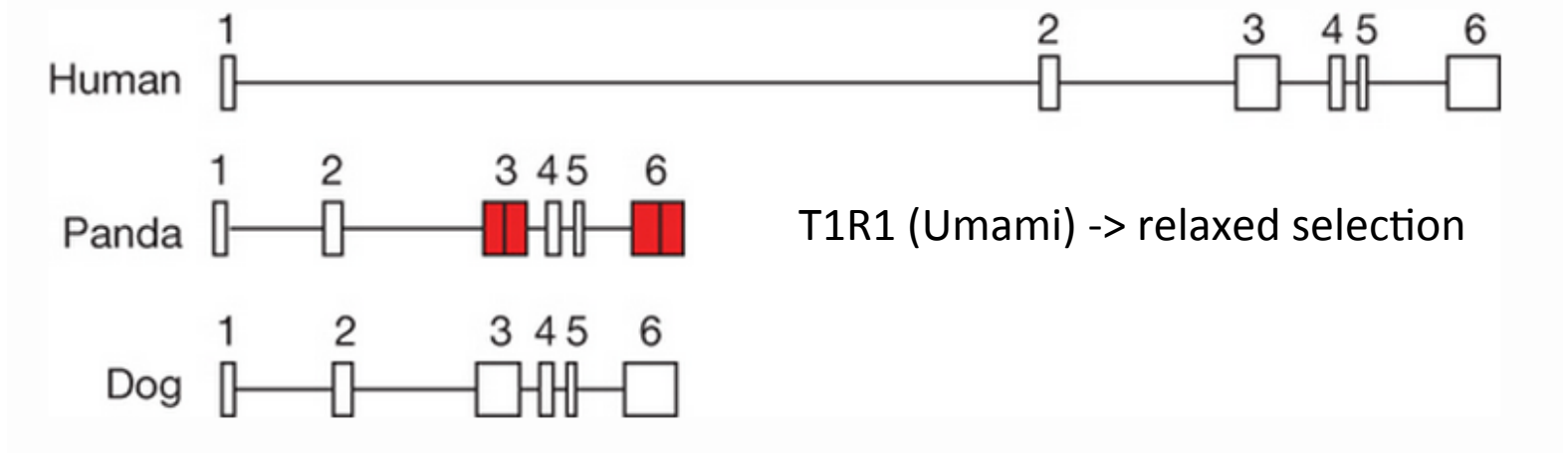


Signatures of Selection in Large Datasets

Pablo Orozco-terWengel

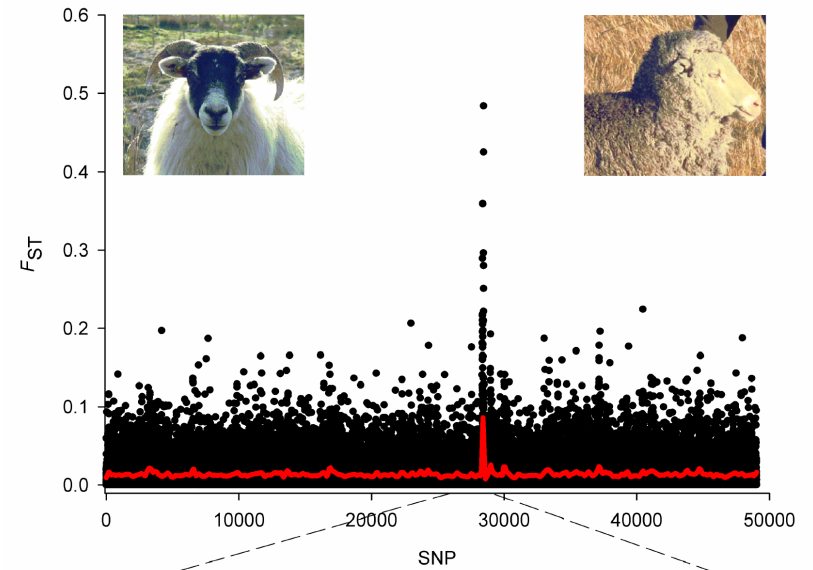
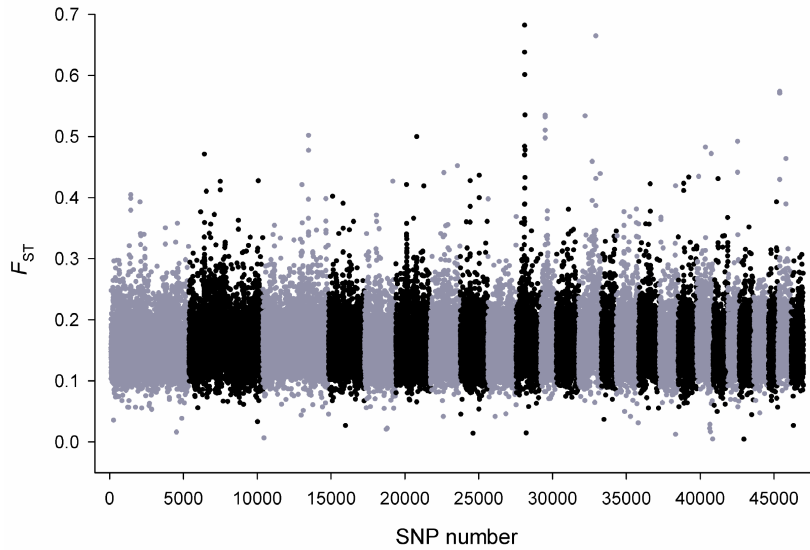
Livestock Genomic Resources in a Changing World
Symposium: The Data Tsunami – how to avoid throwing the baby out with the water!
18th June 2014

Why detect selection?

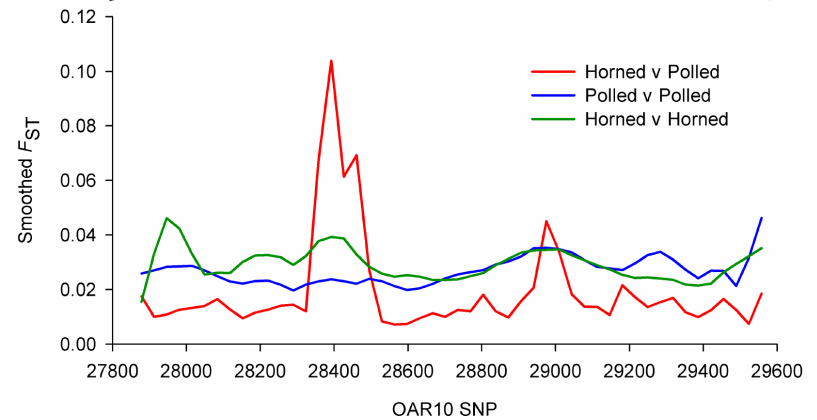


Why detect selection?

Horns or no horns?

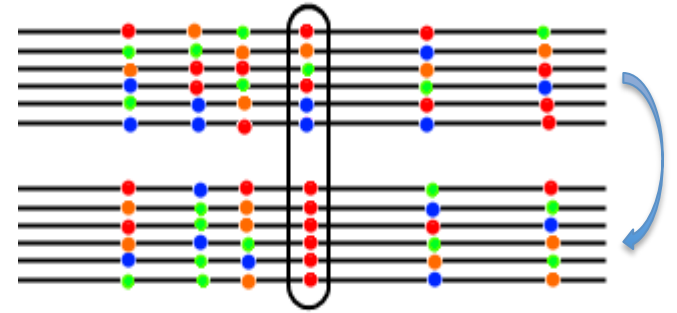
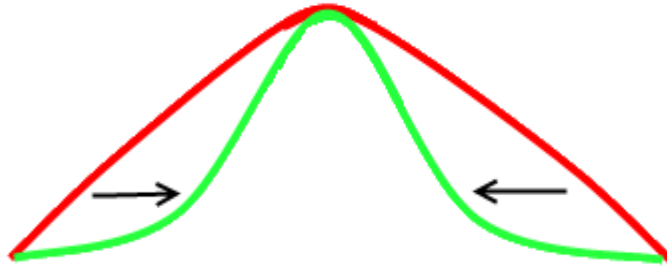


Relaxin/insulin-like (*RXFP2*)



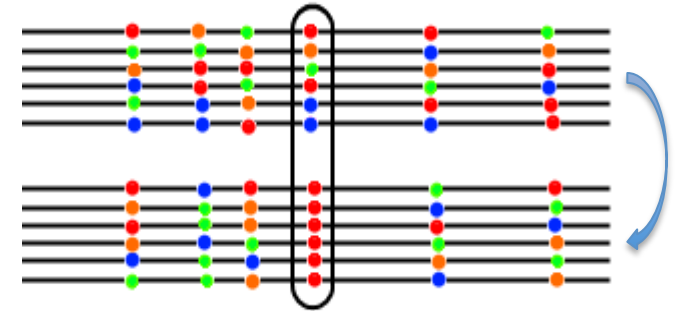
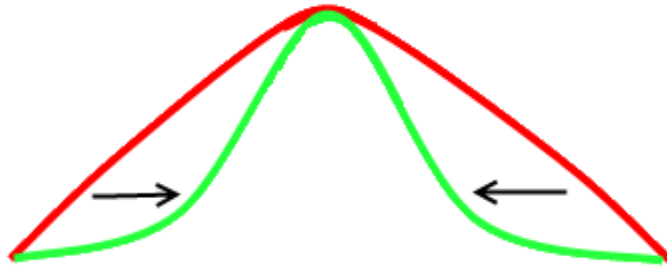
What are we looking for?

negative

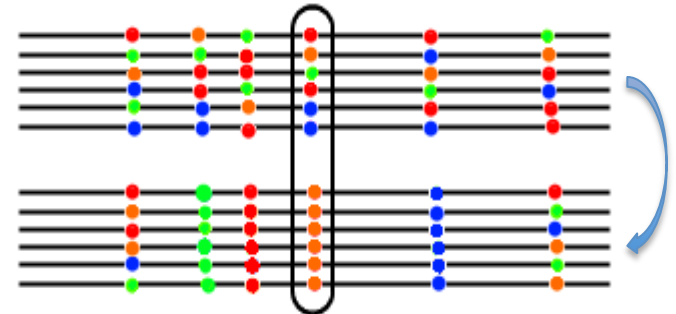
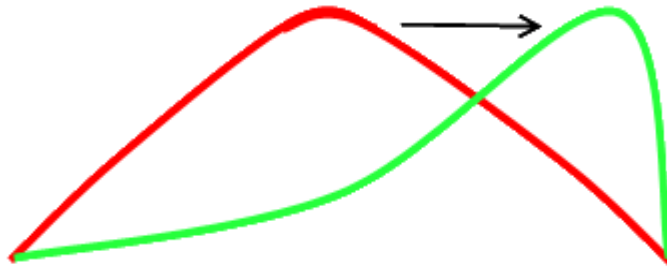


What are we looking for?

negative

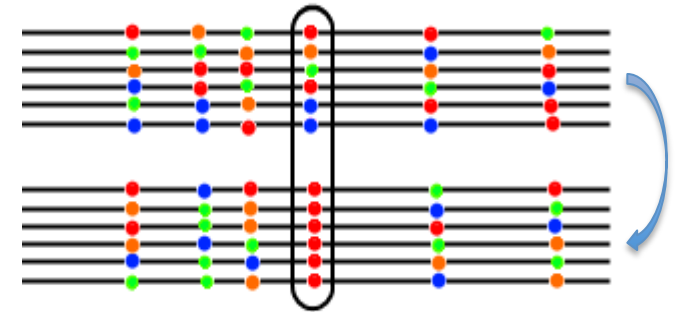
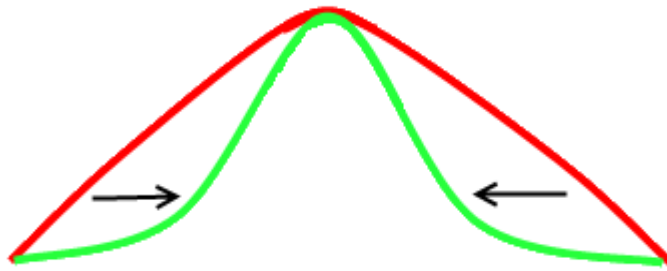


positive

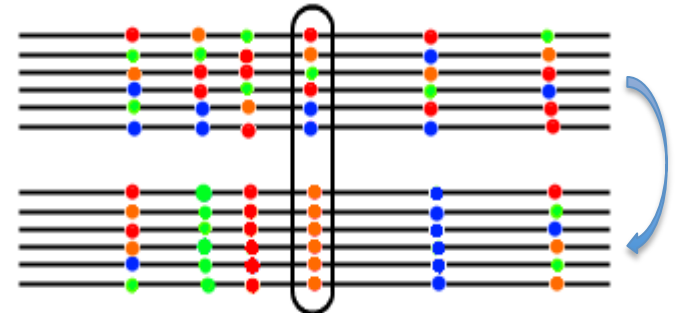
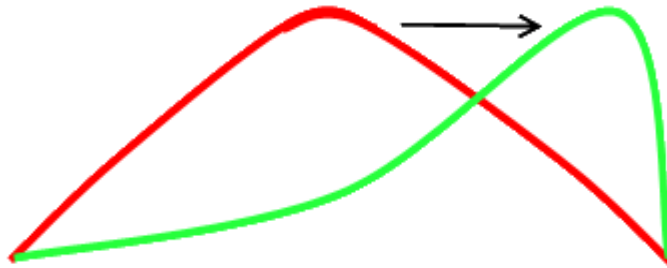


What are we looking for?

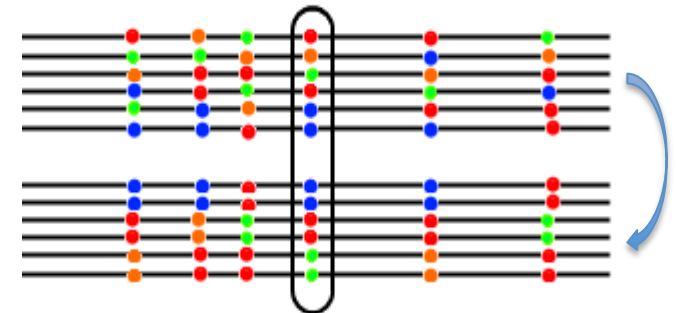
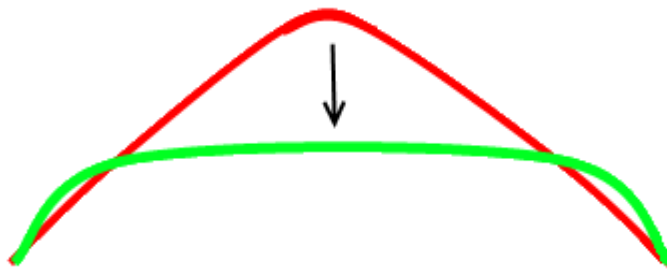
negative



positive

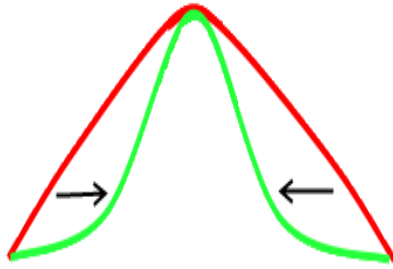


balancing

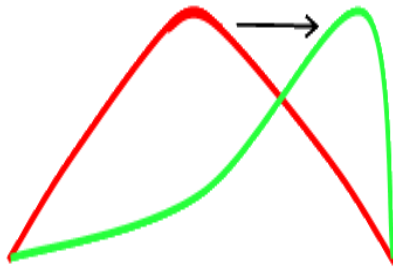


What are we looking for?

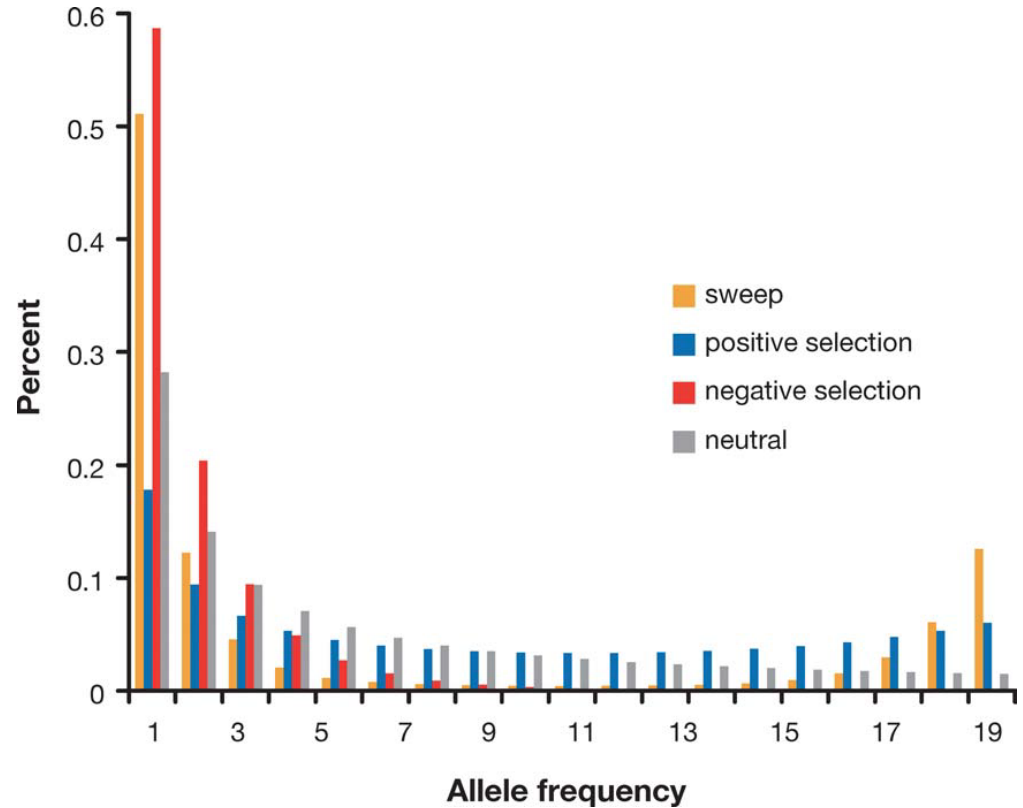
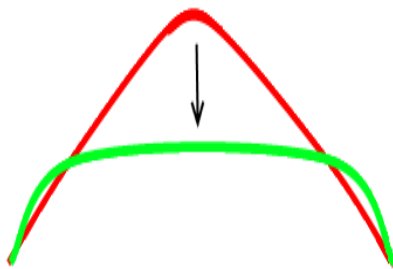
negative



positive



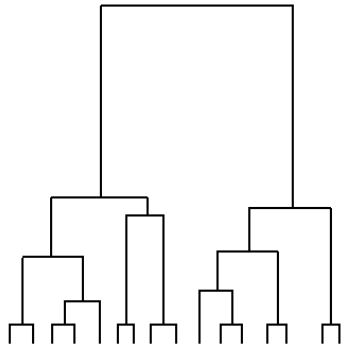
balancing



But....sweeps from standing variation.....

Detection of the signature of selection

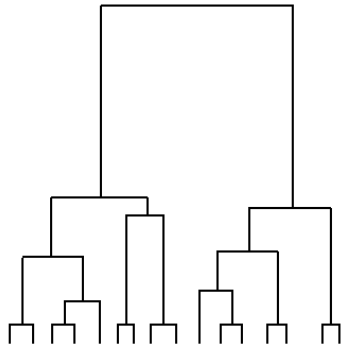
Rate of coalescence in the underlying genealogy



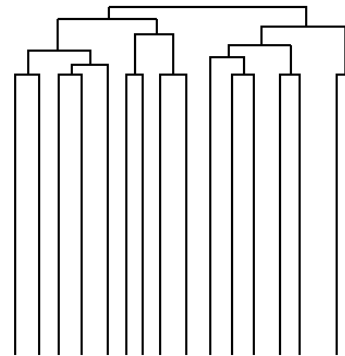
$$P_c(t) = \left(1 - \frac{1}{2N_e}\right)^{t-1} \left(\frac{1}{2N_e}\right)$$

Detection of the signature of selection

Rate of coalescence in the underlying genealogy



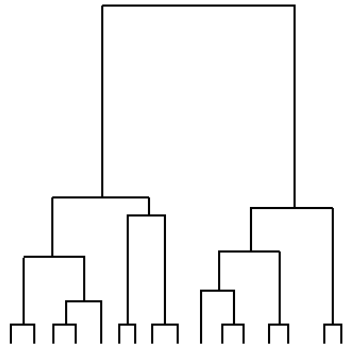
$$P_c(t) = \left(1 - \frac{1}{2N_e}\right)^{t-1} \left(\frac{1}{2N_e}\right)$$



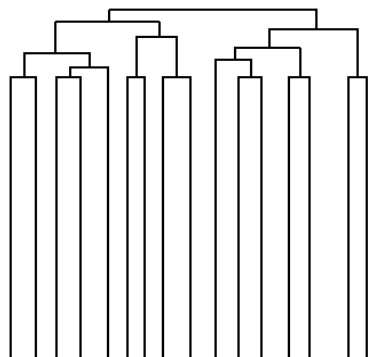
expansion

Detection of the signature of selection

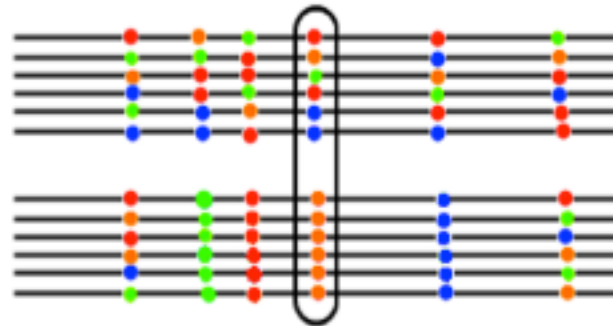
Rate of coalescence in the underlying genealogy



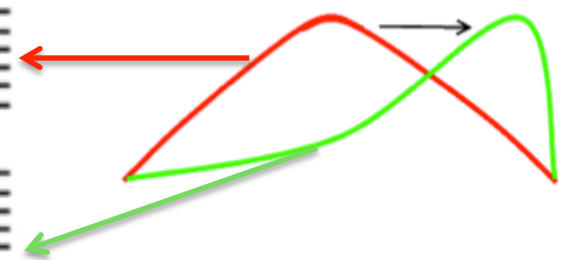
stable



expansion



Positive selection

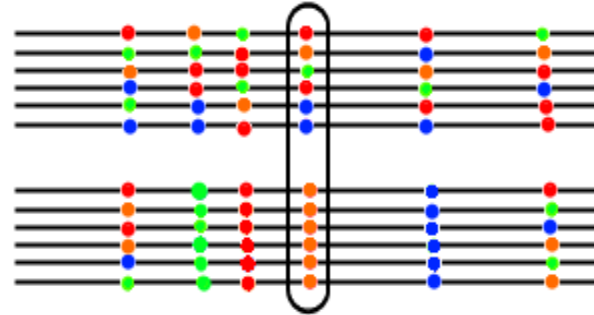


Detection of the signature of selection

$$\Theta_{\pi} = \pi \quad \{1/n \sum \text{differences}/(n * \text{length})\}$$

$$\Theta_S = S \quad \{S/(\sum 1/i)\}$$

$$D = \frac{\hat{\theta}_{\pi} - \hat{\theta}_S}{\sqrt{\text{var}(\hat{\theta}_{\pi} - \hat{\theta}_S)}}$$



$$\Theta_{\pi} \sim \Theta_S ; D \sim 0$$

$$\Theta_{\pi} < \Theta_S ; D < 0$$

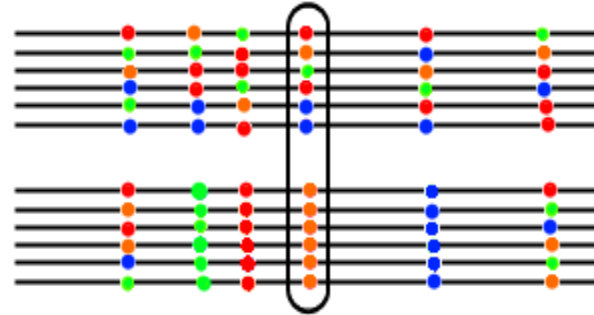
Positive selection

Detection of the signature of selection

$$\Theta_{\pi} = \pi \quad \{1/n \sum \text{differences}/(n * \text{length})\}$$

$$\Theta_S = S \quad \{S/(\sum 1/i)\}$$

$$D = \frac{\hat{\theta}_{\pi} - \hat{\theta}_S}{\sqrt{\text{var}(\hat{\theta}_{\pi} - \hat{\theta}_S)}}$$



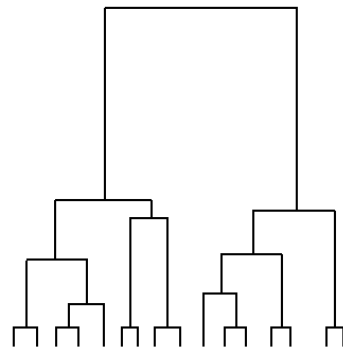
$$\Theta_{\pi} \sim \Theta_S ; D \sim 0$$

$$\Theta_{\pi} < \Theta_S ; D < 0$$

Positive selection

Theta and the coalescent

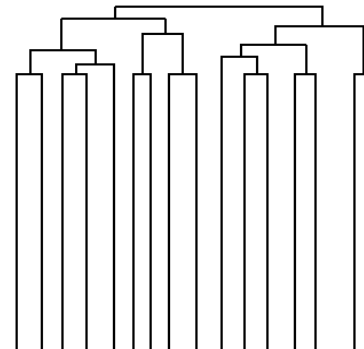
$$\Theta = 4Ne\mu$$



$$\Theta_{\pi} \sim \Theta_S$$

$$D \sim 0$$

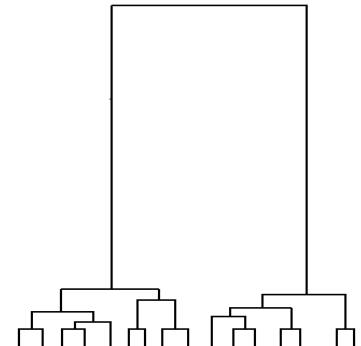
Equilibrium



$$\Theta_{\pi} < \Theta_S$$

$$D < 0$$

Positive Sel./Expansion

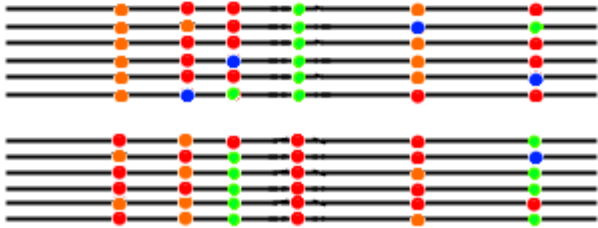


$$\Theta_{\pi} > \Theta_S$$

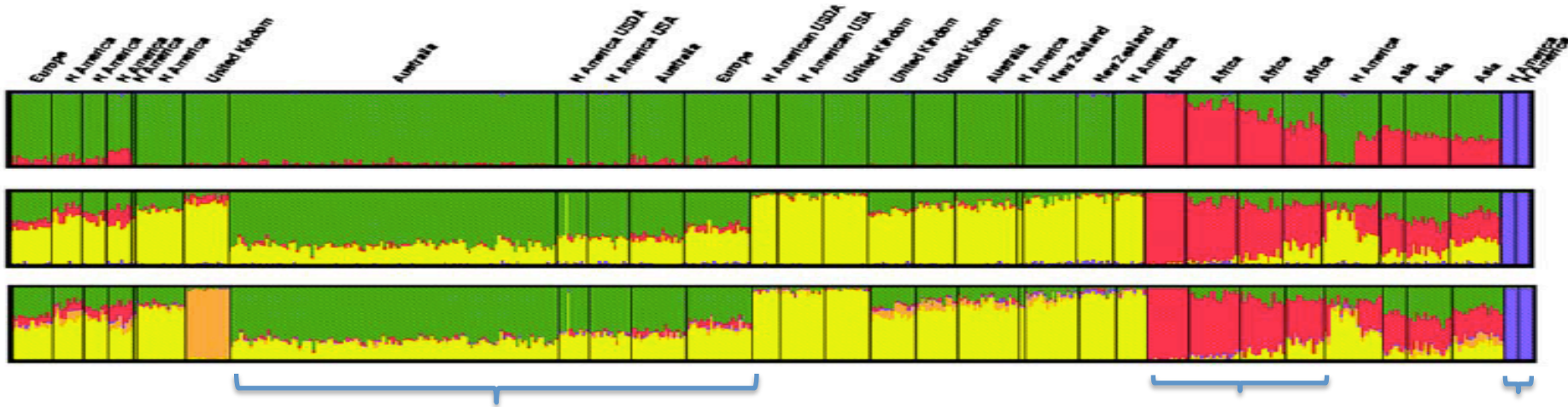
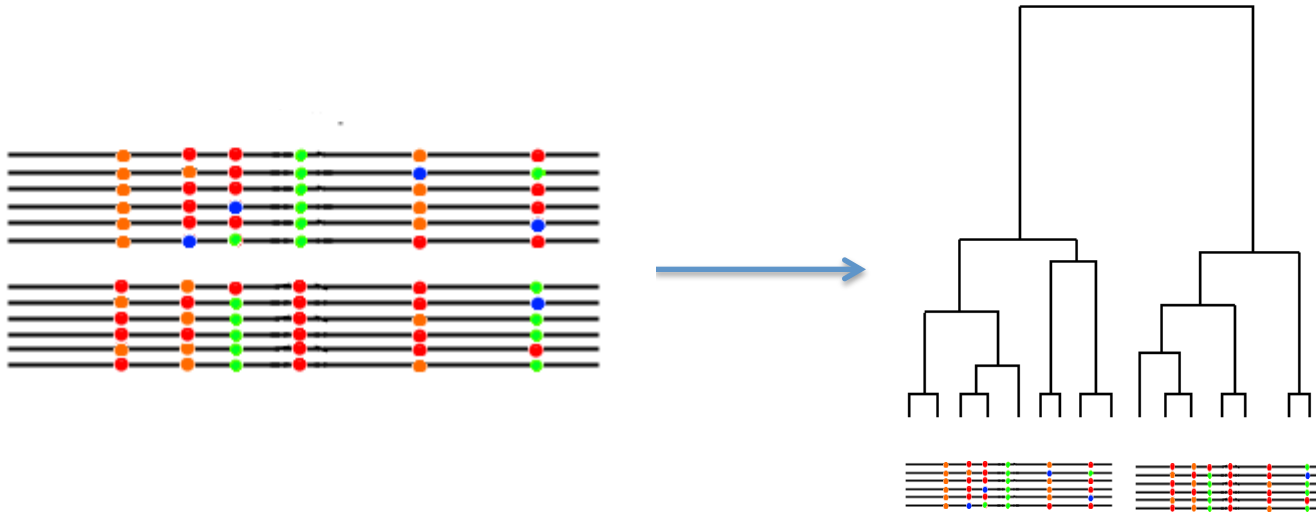
$$D > 0$$

Balancing Sel./Deep Div.

Detection of the signature of selection



Detection of the signature of selection



Definition of units for downstream analysis

Detection of the signature of selection

plenty of methods available:

DISTRIBUTION OF GENE FREQUENCY AS A TEST OF THE THEORY OF THE SELECTIVE NEUTRALITY OF POLYMORPHISMS^{1,2}

LI WENTON AND JESSE KRAKAUER
Department of Biology
60637

Copyright © 1989 by the Genetics Society of America

Statistical Method for Testing the Neutral Mutation Hypothesis of Polymorphism

Fumio Tajima
Department of Biology, Kyushu University
Manuscript received August 7, 1989
Accepted for publication March 20, 1990

1,2, Japan

Statistical Tests of Neutrality of Mutations

Yun-Xin Fu¹ and Wen-Hsiung Li

Center for Demographic and Population Genetics, University of Texas, Austin, Texas 78712

Manuscript received July 10, 1997
Revised copy accepted March 20, 2000

Copyright © 2000 by the Genetics Society of America

Hitchhiking Under Positive Darwinian Selection

Justin C. Fay* and Chung-I Wu*,†

*Committee on Genetics and †Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637
Manuscript received August 7, 1999
Accepted for publication March 20, 2000

LETTERS

Whole-genome resequencing reveals loci under selection during chicken domestication

Carl-Johan Rubin^{1*}, Michael C. Zody^{1,2*}, Jonas Eriksson¹, Jennifer R. S. Meadows¹, Ellen Sherwood³, Matthew T. Webster¹, Lin J. Orjan Carlborg⁶, Bertrand F. Meyer⁴, Axel Ingman⁵, Ted Sharpe², Sojeong Ka⁵, Finn Hallböök⁵, Francois Besnier⁶, & Leif Andersson^{1,6}
¹Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; ²Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; ³Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; ⁴Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; ⁵Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; ⁶Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

A Map of Recent Positive Selection in the Human Genome

Benjamin F. Voight⁶, Sridhar Kudaravalli⁶, Xiaoquan Wen, Jonathan K. Pritchard¹
Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

The identification of signals of very recent positive selection provides information about the adaptation of modern humans to local conditions. We report here on a genome-wide scan for signals of very recent positive selection in favor of variants that have not yet reached fixation. We describe a new analytical method for scanning single nucleotide

ABSTRACT

Positive selection can be inferred from its effect on linked neutral variation. In the restrictive case when there is no recombination, all linked variation is removed. If recombination is present but rare, both deterministic and stochastic models of positive selection show that linked variation hitchhikes to either low or high frequencies. While the frequency of variation can be influenced by a number of

A Microsatellite Variability Screen for Positive Selection Associated With the "Out of Africa" Habitat Expansion of *Drosophila melanogaster*

M. O. Kauer,¹ D. Dieringer¹ and C. Schlötterer²
Institut für Tierzucht und Genetik, 1210 Wien, Austria
Manuscript received October 9, 2002
Accepted for publication March 20, 2003

selScan: an efficient multi-threaded program to perform EHH-based scans for positive selection

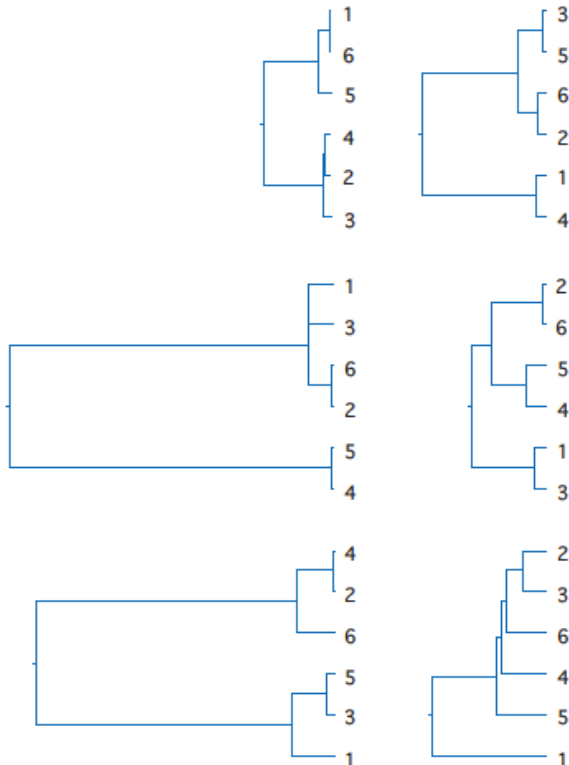
Zachary A. Szpiech* and Ryan D. Hernandez

May 20, 2014

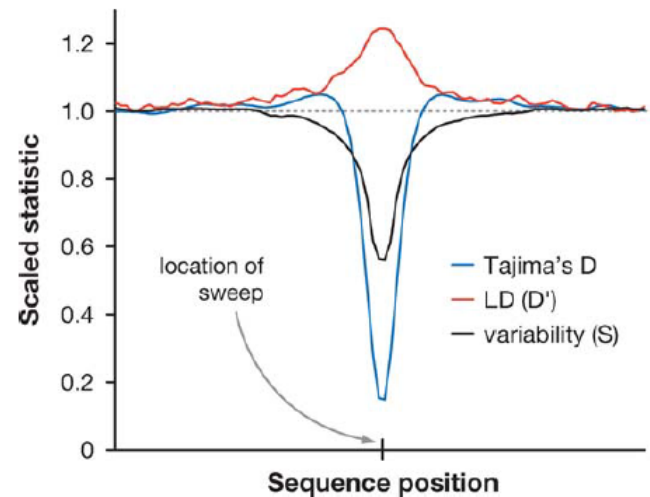
Detection of the signature of selection

The value of haplotypes vs. single SNPs

Single genealogies have a random component



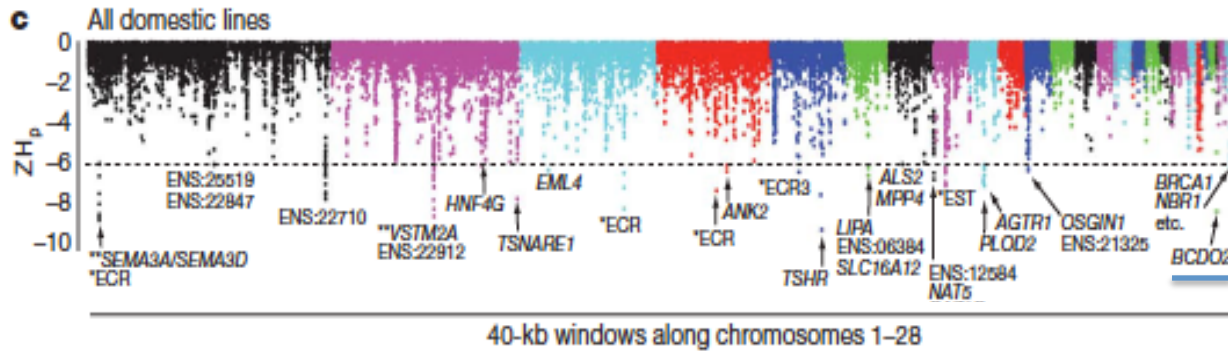
Haplotypes carry information on stretches of correlated points



Detection of the signature of selection

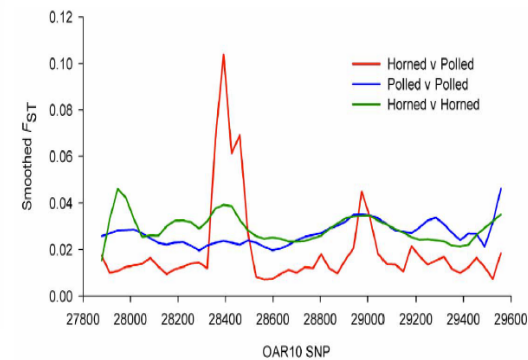
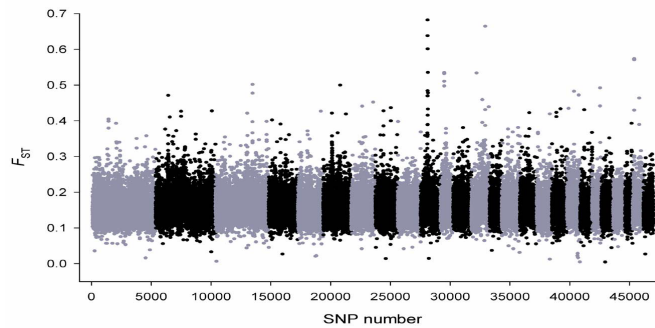
In the absence of haplotypes

ZHp



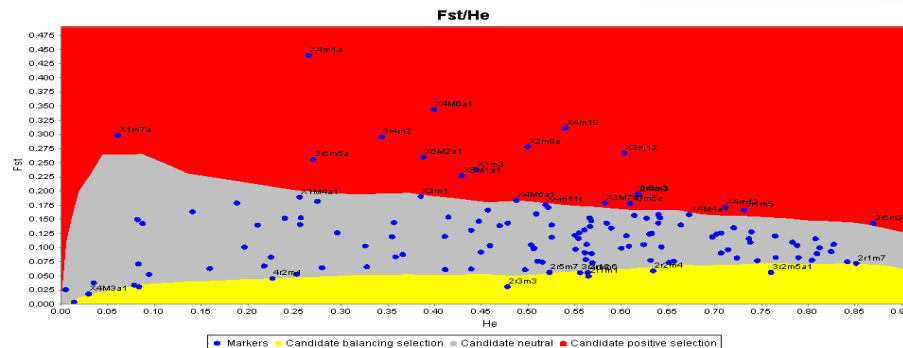
Rubin et al 2010 Nature

Fst outliers



Kijas et al 2012 PLoS Biology

Fst/H outliers



Nichols & Beaumont 1996

(Orozco-terWengel et al unpublished)

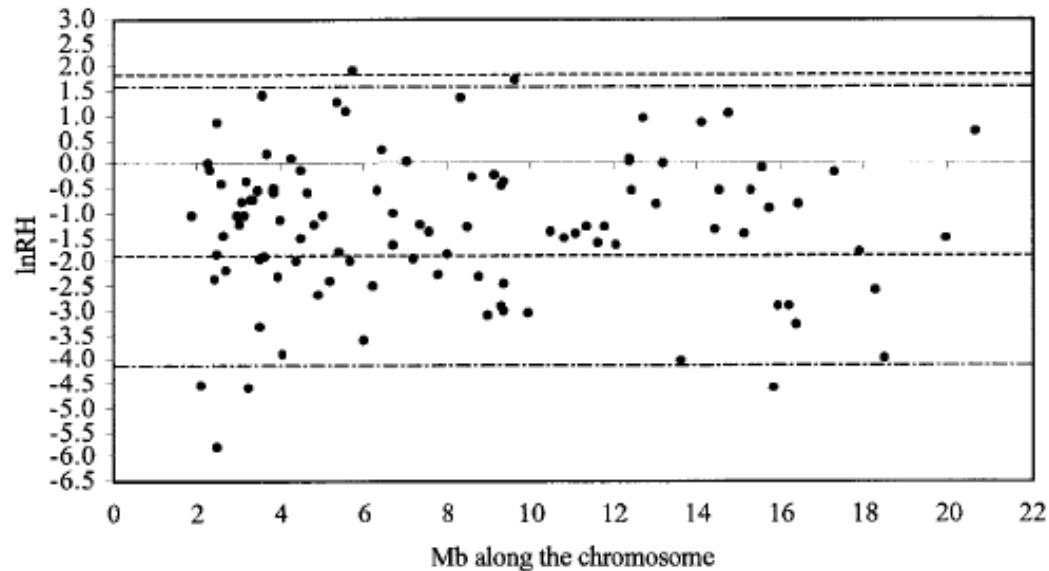
Detection of the signature of selection

Making use of ancestral versus derived information

$$\ln[E(RH)] = \ln \left[E \left(\frac{\theta_{\text{Pop1}}}{\theta_{\text{Pop2}}} \right) \right]$$

θ_{Pop1} derived population

θ_{Pop2} ancestral population



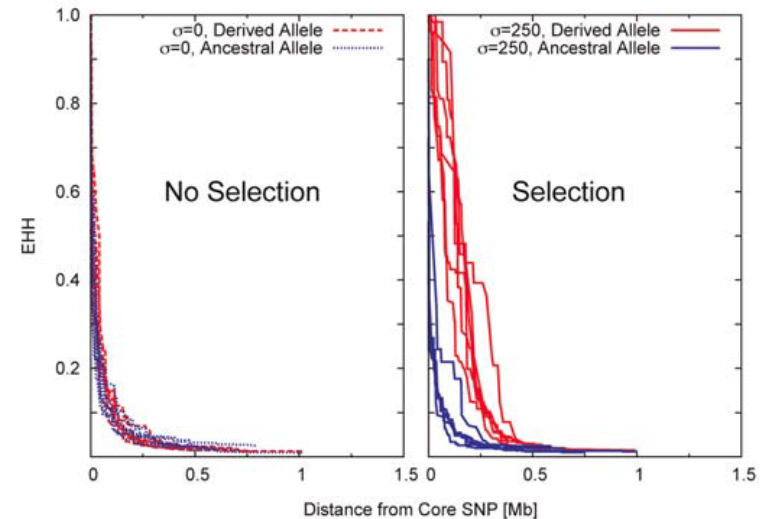
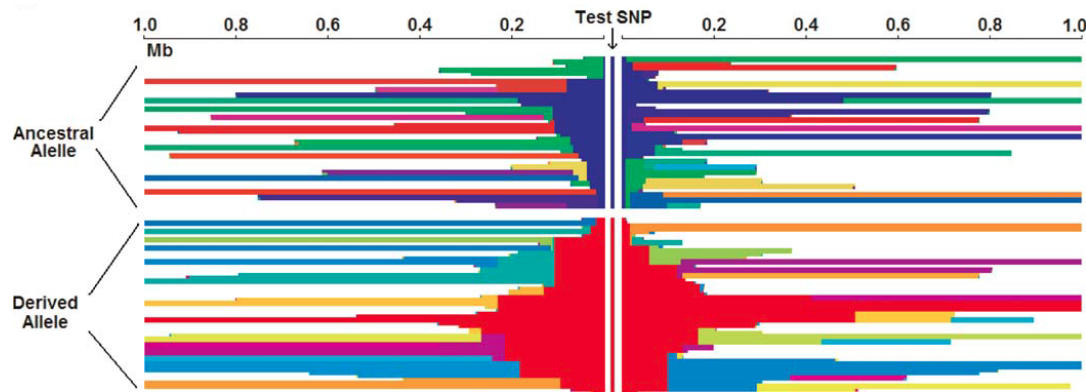
Detection of the signature of selection

Making use of haplotypes

$$iHS = \frac{\ln\left(\frac{iHH_A}{iHH_D}\right) - E_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}{SD_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}$$

ancestral allele

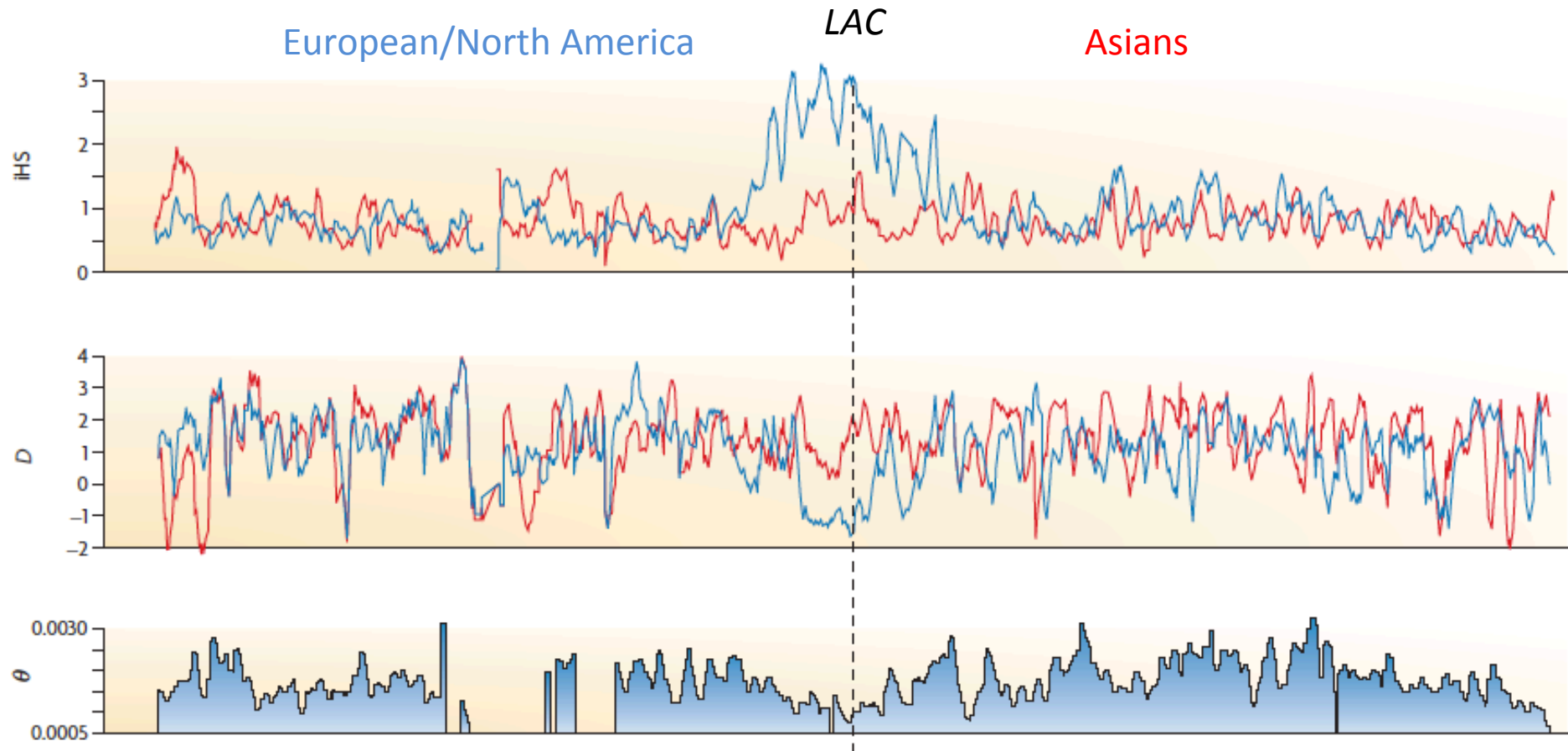
derived allele

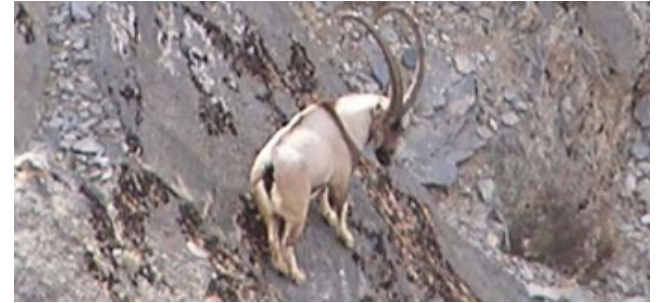


Allow detection of “very recent and incomplete” sweeps

Detection of the signature of selection

Making use of haplotypes





General Info

Species	Origin	Code	Sample size
<i>O. orientalis</i>	Iran	IROO	16
<i>O. aries</i>	Iran	IROA	20
	Morocco	MOOA	20
	Industrial breeds	indusOA	20
<i>C. aegagrus</i>	Iran	IRCA	19
<i>C. hircus</i>	Iran	IRCH	20
	Morocco	MOCH	20
	Industrial breeds	indusCH	10

~36M SNPs

~24M SNPs

General Info

Species	Origin	Code	Sample size
<i>O. orientalis</i>	Iran	IROO	16
<i>O. aries</i>	Iran	IROA	20
	Morocco	MOOA	20
	Industrial breeds	indusOA	20
<i>C. aegagrus</i>	Iran	IRCA	19
<i>C. hircus</i>	Iran	IRCH	20
	Morocco	MOCH	20
	Industrial breeds	indusCH	10

~36M SNPs

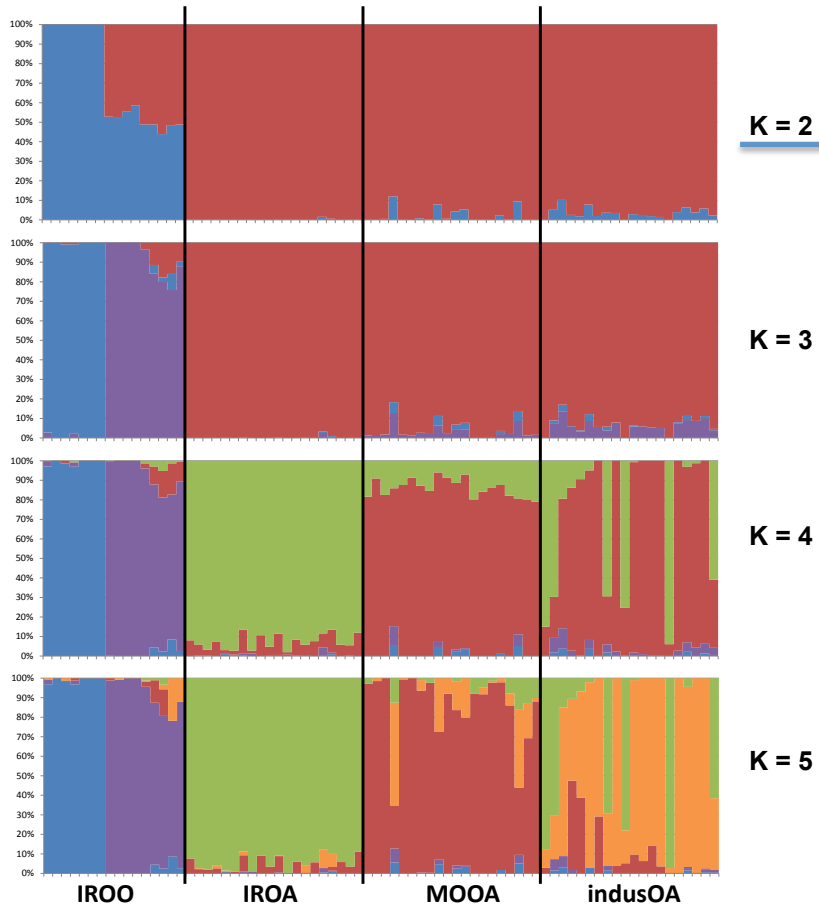
~24M SNPs

Fst

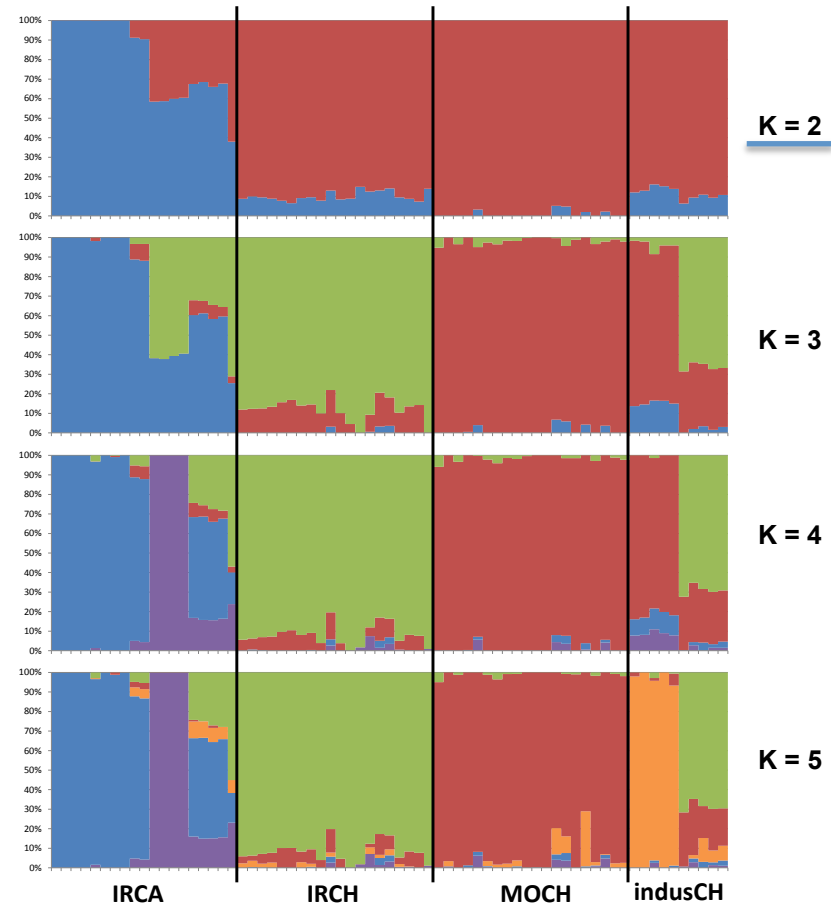
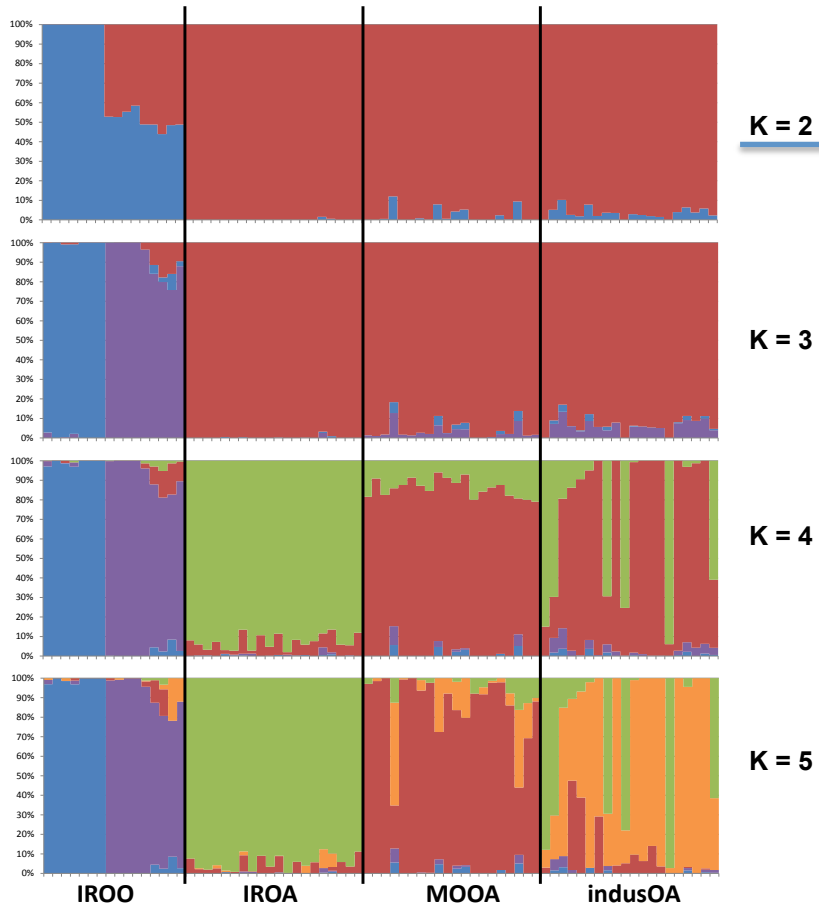
	IROA	indusOA	IROO
MOOA	0.017	0.007	0.087
IROA	-	0.011	0.088
indusOA	-	-	0.080

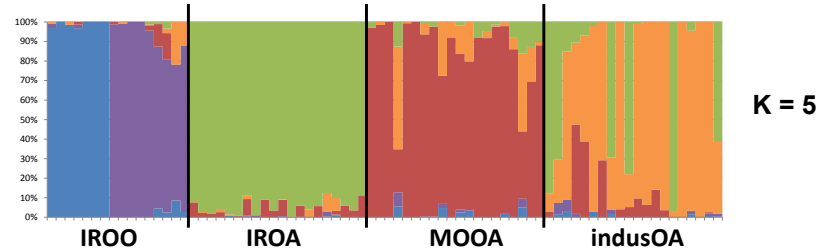
	IRCH	indusCH	IRCA
MOCH	0.031	0.031	0.063
IRCH	-	0.021	0.045
indusCH	-	-	0.058

Structure Analysis

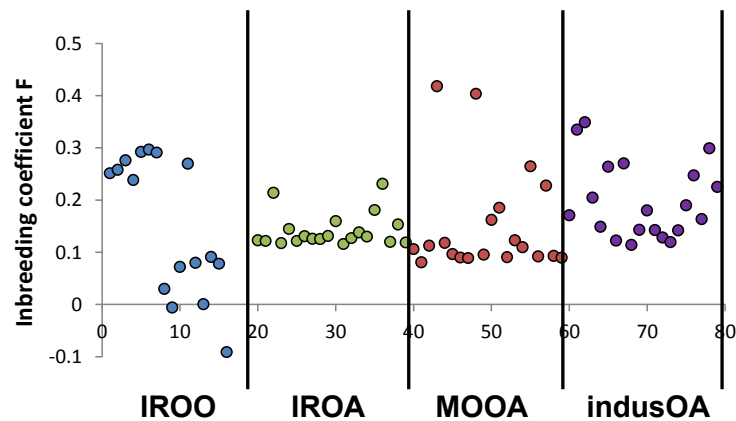


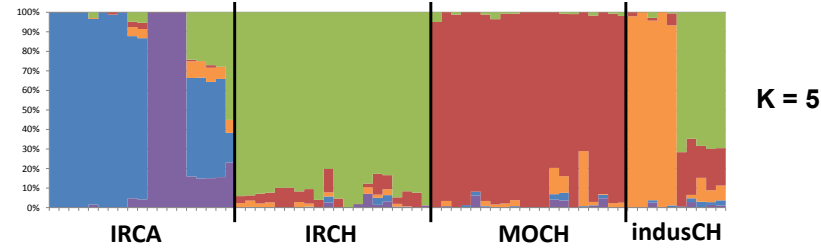
Structure Analysis



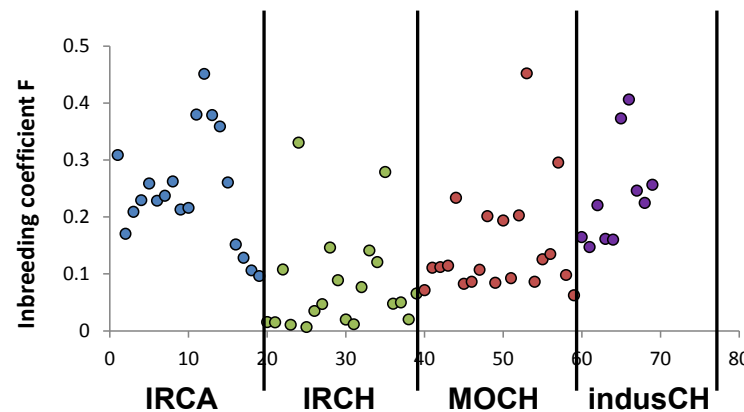


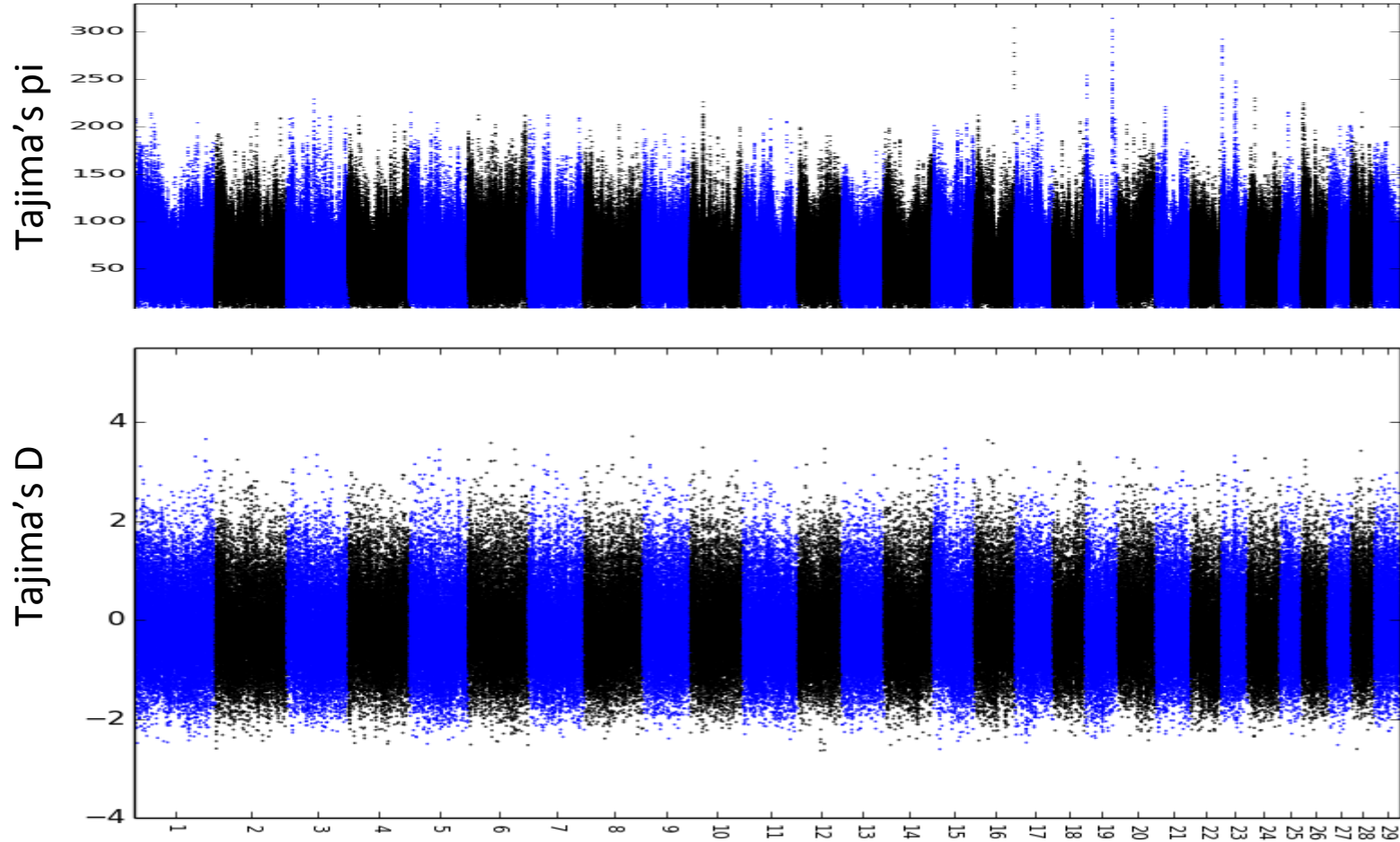
	IROA	MOOA	indusOA	IROO	IROO-1	IROO-2
Sample size	20	20	20	16	9	7
Nb. SNPs	25.6 M	27.2 M	25.1 M	30.1 M	28.5 M	14.3 M
mean pi	0.139	0.145	0.141	<u>0.210</u>	0.171	0.114
mean F	0.15	0.16	0.20	0.20	0.06	0.27



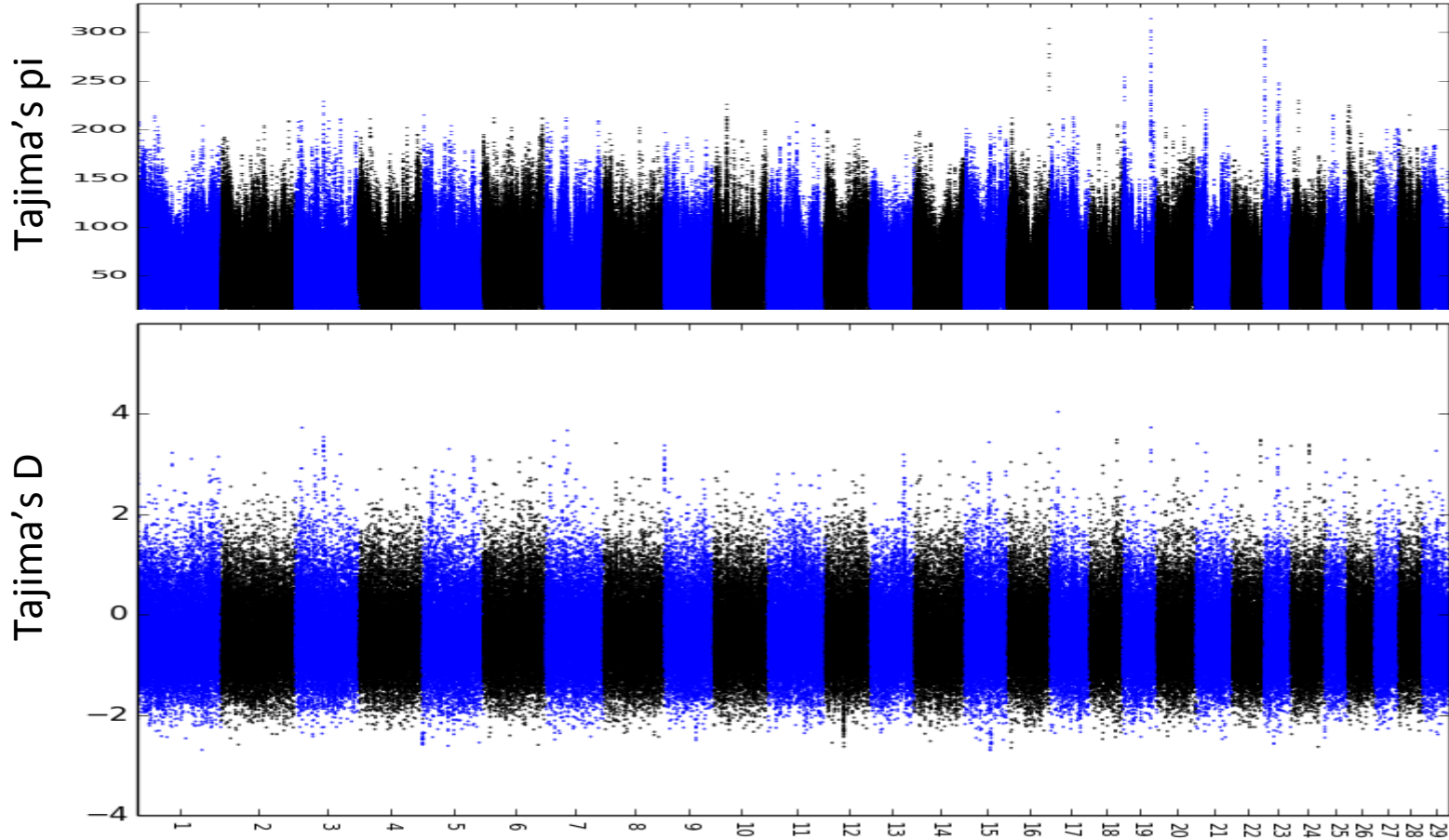


	IRCH	MOCH	indusCH	IRCA	IRCA-1	IRCA-2	IRCA-3
Sample size	20	20	10	19	10	4	5
Nb. SNPs	21.7	19.2	11.2	17.4	11.4	6.5	12.9
mean pi	0.125	0.118	0.092	<u>0.109</u>	0.099	0.077	0.121
mean F	0.08	0.15	0.24	0.24	0.23	0.39	0.15

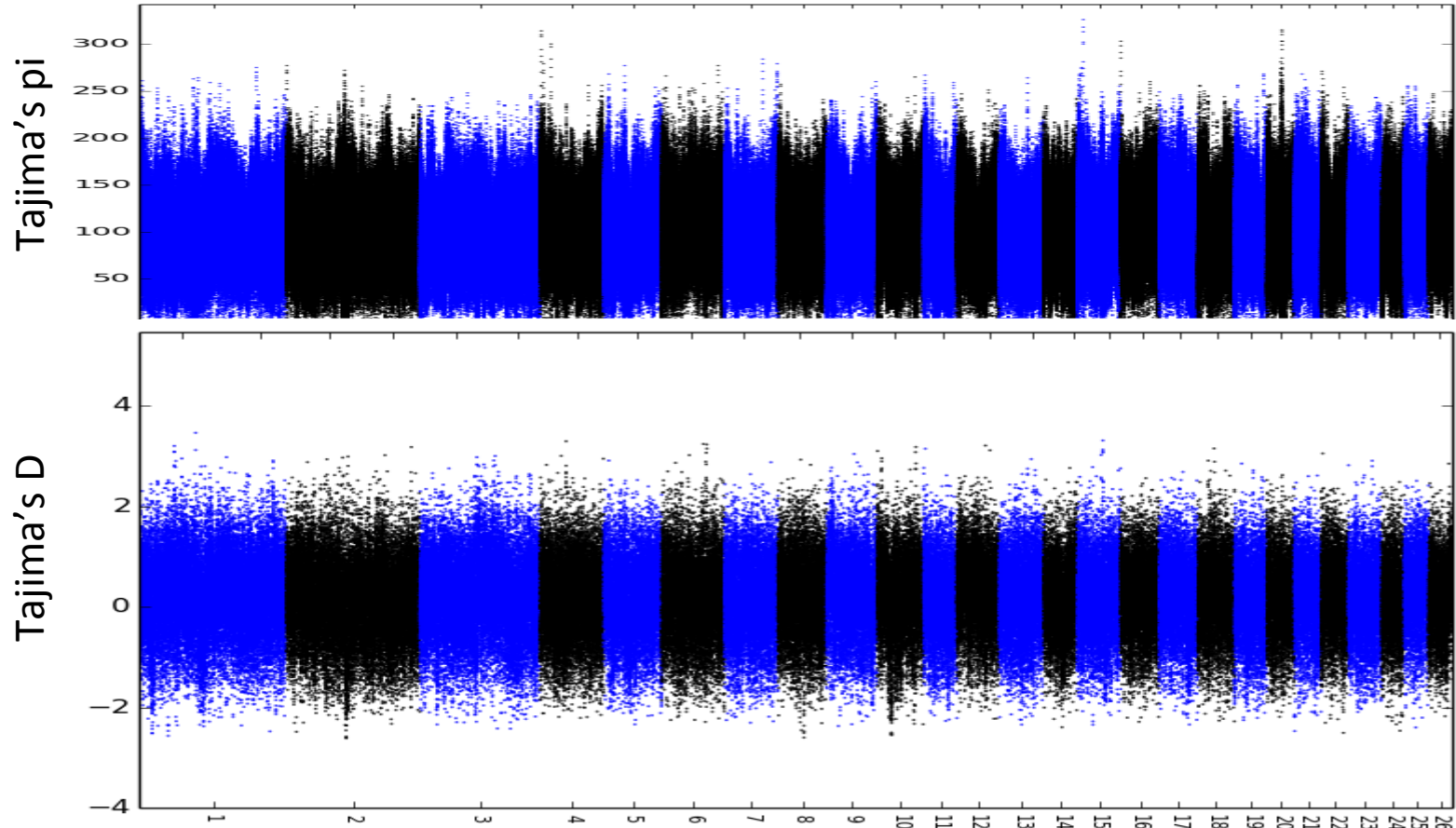




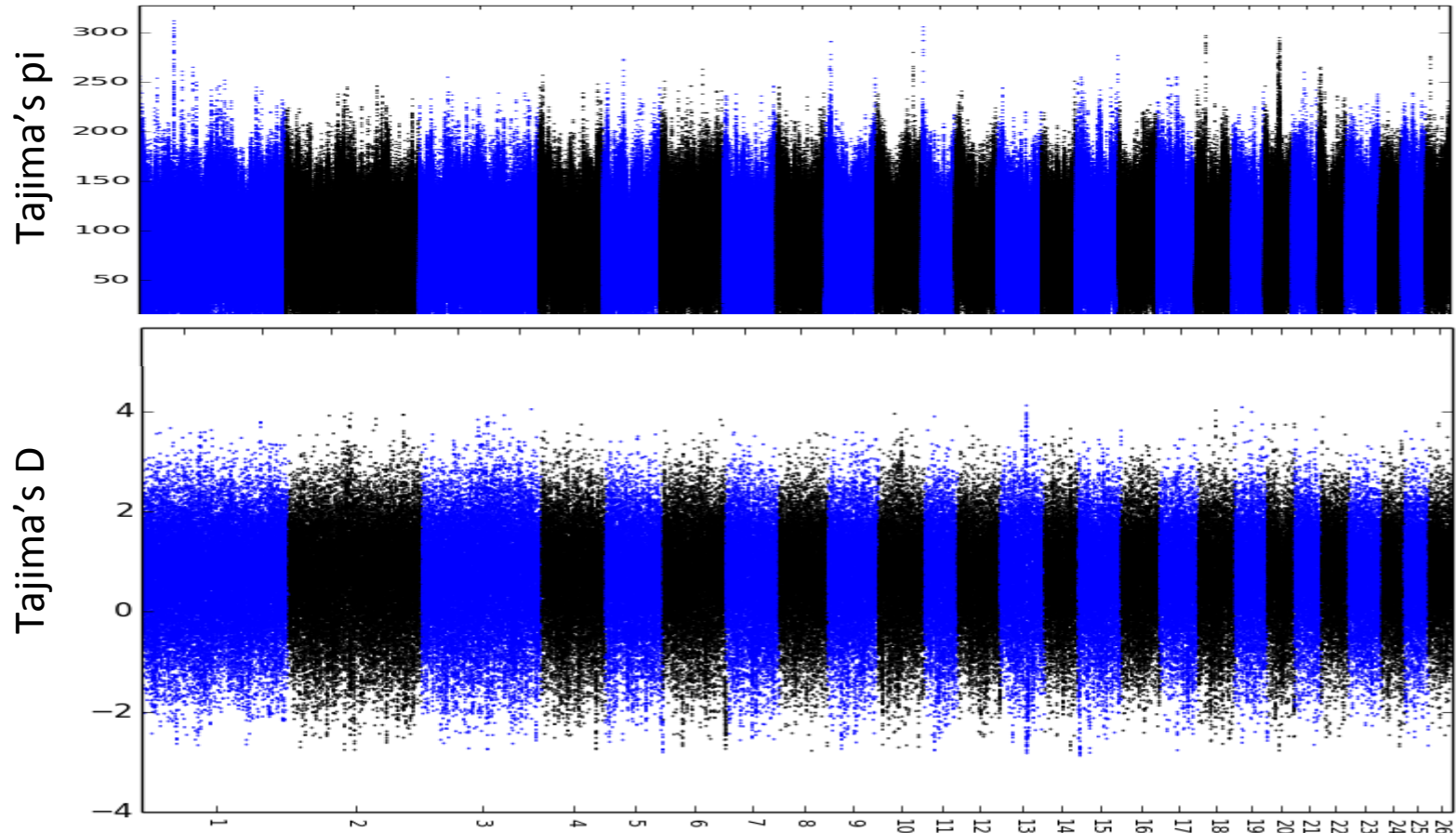
Average Tajima's D -0.1414675 (± 0.8086764)
Less than 1% SNPs Tajima's D ≤ -2



Average Tajima's D -0.3702259 (± 0.7240281)
Less than 1% SNPs Tajima's D ≤ -2
Sig high $p < 2.2e-16$



Average Tajima's D 0.1878116 (+/- 0.7619021)
Less than 1% SNPs Tajima's D ≤ -2

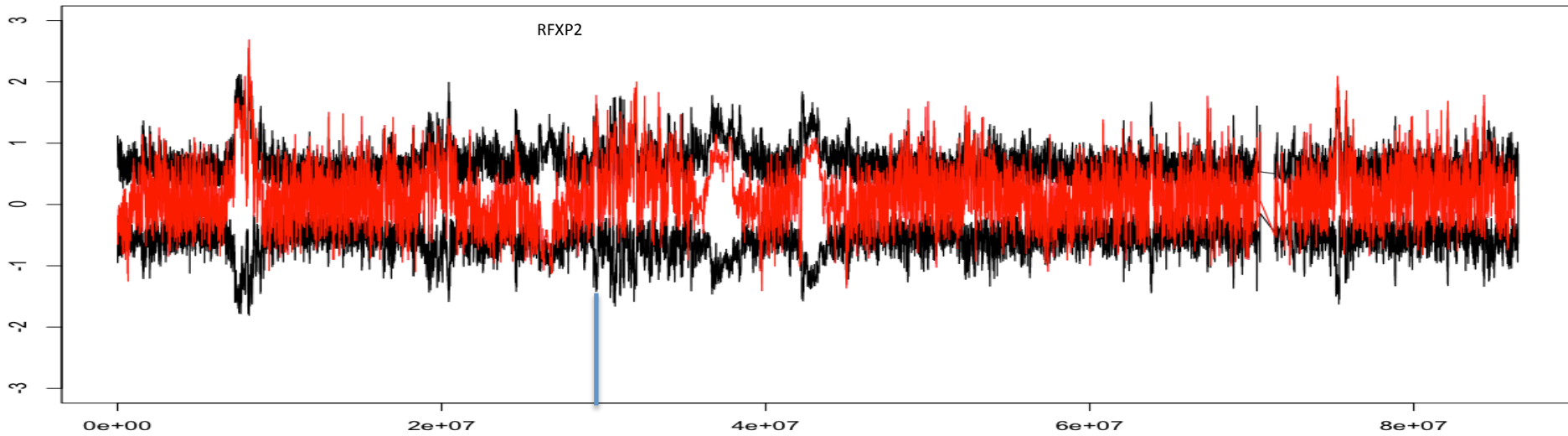


Average Tajima's D 0.6455212 (+/- 0.9450472)
Less than 1% SNPs Tajima's D ≤ -2
Sig high $p < 2.2e-16$

XP-EHH



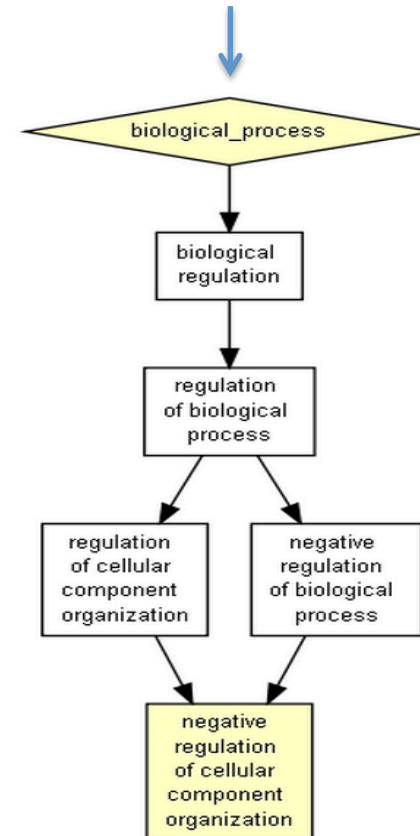
50.000 SNPs OAR10



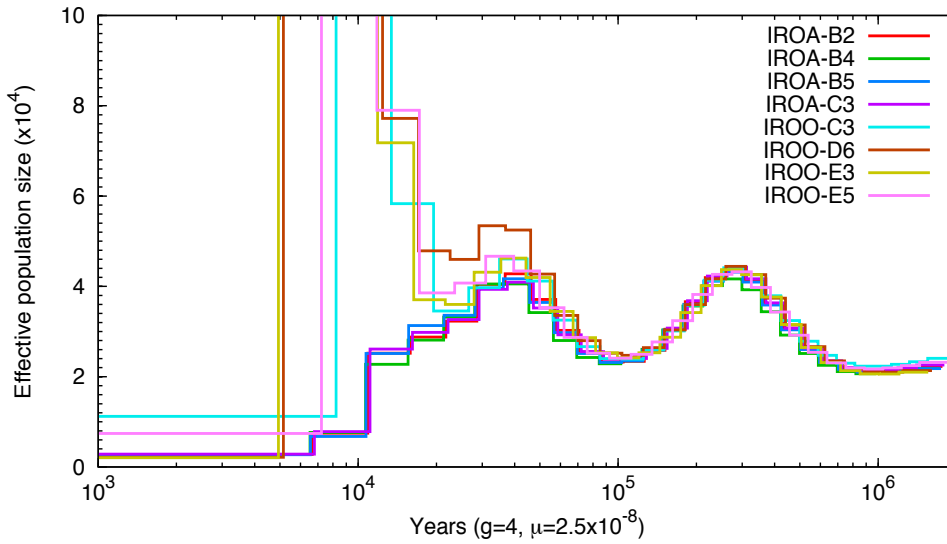
	Mouflon	Sheep
Average Taj D	0.1878116 (+/- 0.8)	0.6455212 (+/- 0.9)
All Genes	199	484
Char. Genes	39	103

Genes Common to Mouflon and Sheep

Gene	Process
AHSA2	Activator of HSP90
EHD1	Encocytosis and platelet activation
KCTD17	K channel related to redox protection pathways

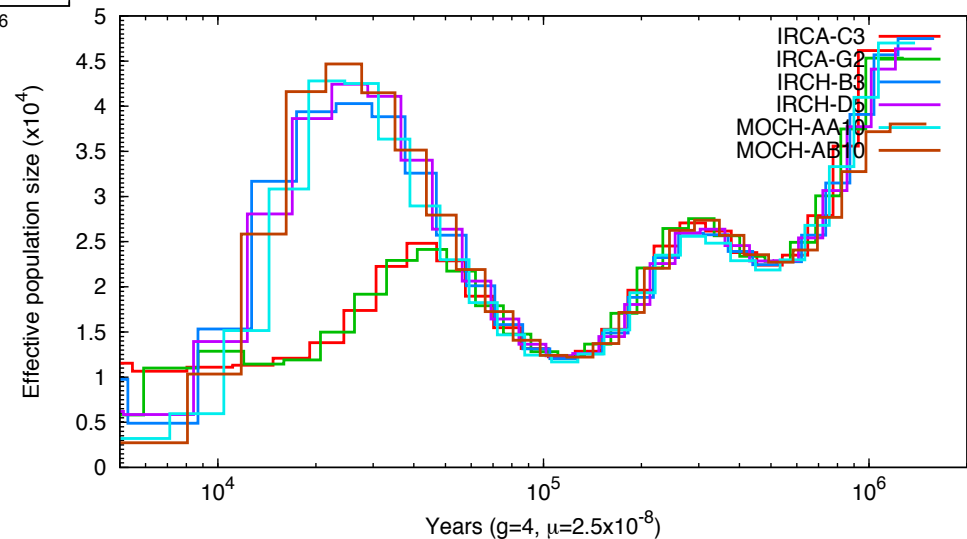
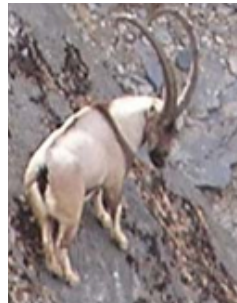


But, ideally, a method for selection would allow to account for N_e changes!!



Iranian mouflon and sheep

Iranian bezoar and goats



Acknowledgments

M. Bruford J. Kijas F. Pompanon
P. Taberlet S. Joost S. Stuckie
A. Stella F. Biscarini F. Alberti
C. Mesembe I. Streeter F. Boyer
B. Benjelloun W. Zamani L. Colli
P. Ajmone-Marsan and NextGen Consortium

Mario Barbato

