

# Estimating the potential of wild relatives and indigenous breeds to act as genomic resources: the case of sheep and goats

François Pompanon  
on behalf of the Nextgen Consortium





# Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options

Coordinator: P. Taberlet – CNRS France  
2010 – 2014



Padano, Italy



Teramo, Italy



Errachidia, Morocco



Grenoble, France



Lausanne, Switzerland



Gorgan, Iran



Kampala, Uganda



Cardiff, UK



Brisbane, Australia



Piacenza, Italy



Cambridge, UK



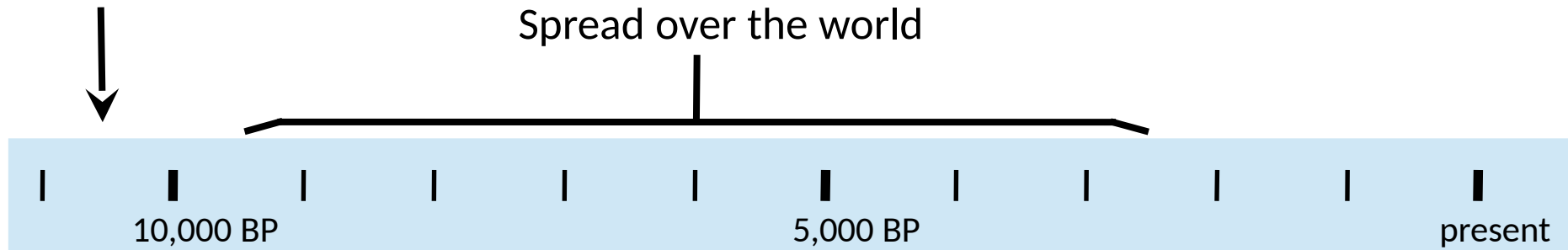
Paris, France

# Are cattle, sheep, and goats endangered species?

P. TABERLET,\* A. VALENTINI,† H. R. REZAEI,\*‡ S. NADERI,\*§ F. POMPANON,\* R. NEGRINI¶  
and P. AJMONE-MARSAN¶\*\*

Domestication of cattle, sheep,  
and goats in the Middle East

Spread over the world



Traditional breeding of local populations  
sustainable management and local adaptation

breed concept

Beginning of  
intensive selection



← Corsican goats



Saanen →

## **Threats on industrial breeds ...**

low number of founders and ancestors, inbreeding and selection:

strong decrease of genetic diversity ( $N_e \approx 50$  for Holstein cattle, 150 for Saanen goats, ....)

Genetic diseases

Rapid decline in fertility

## **and traditional breeds ...**

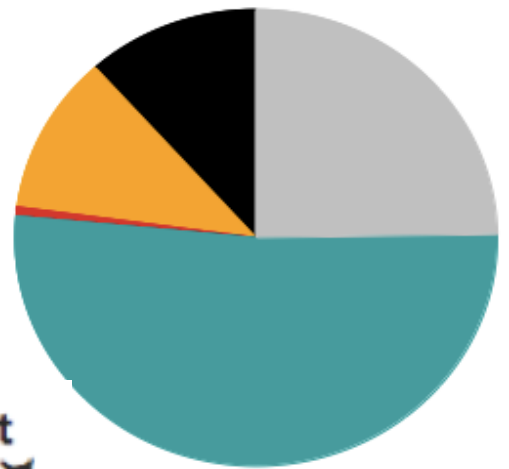
Extinction

Introgression from highly productive breeds

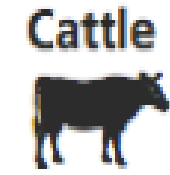
# Breeds status (all breeds) and Population size (billion)



1.4



-  extinct
-  endangered
-  critical
-  not at risk
-  unknown



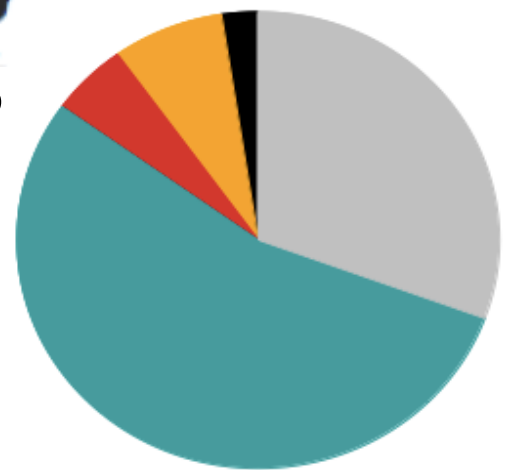
35%



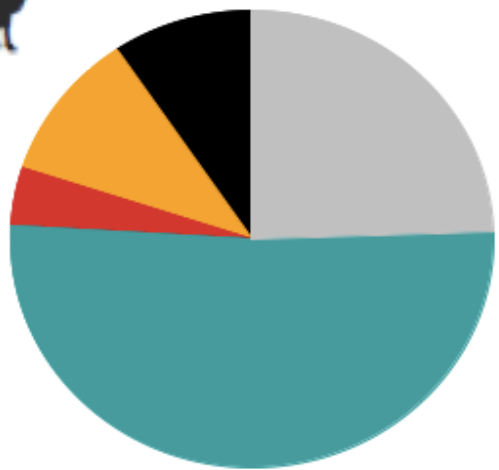
18%



0.9



1.1



30%

threatened + extinct  
among local breeds

## NGS for assessing AnGR

1- Assess the performance of surrogate marker systems compared with whole genome sequence data for preserving biodiversity

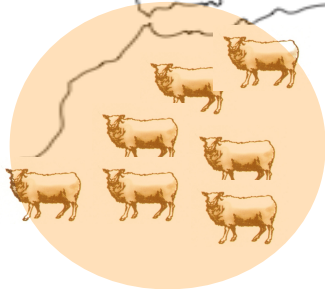
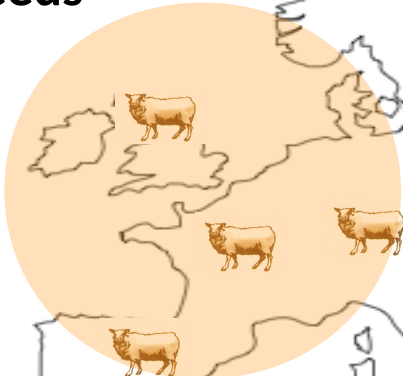
2- Assess the value of wild ancestors and indigenous breeds as genomic resources

3- Consider adaptive genes (local adaptation) for defining genomic resources

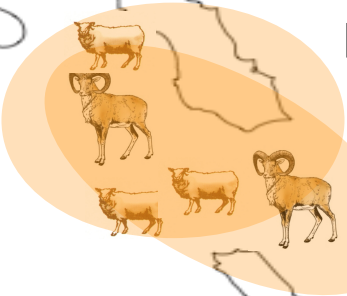
# Sampling design



Industrial breeds  
(n~15-75)



Local breeds (n~160)



Local breeds (n~20)

Wilds (n~20)



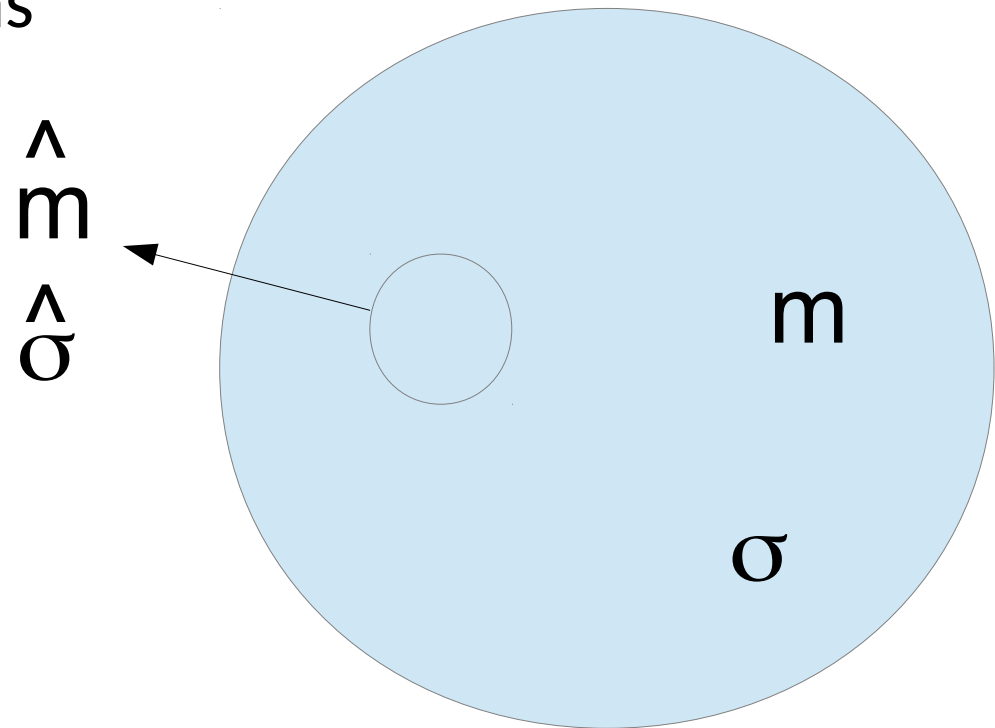
WGS @ 10 X (Illumina HighSeq<sup>®</sup>)



# 1- Performance of surrogate marker systems

How sampling markers in the genome ?

Precise and unbiased estimations



Microsatellites, CNVs,

SNPs, AFLPs & RAD-Seq like, GBS, ESTs & exome, ...

Neutral & selected markers



# Moroccan population & wild relatives

## Filtered phased WGS datasets

Sheep



Goat



162 ids

43.5M variants



*O. aries*

WGS dataset n=30

14 ids

29.3M variants



*O. orientalis*

WGS dataset n=14

163 ids

31.8M variants



*C. hircus*

WGS dataset n=30

18 ids

17.4M variants



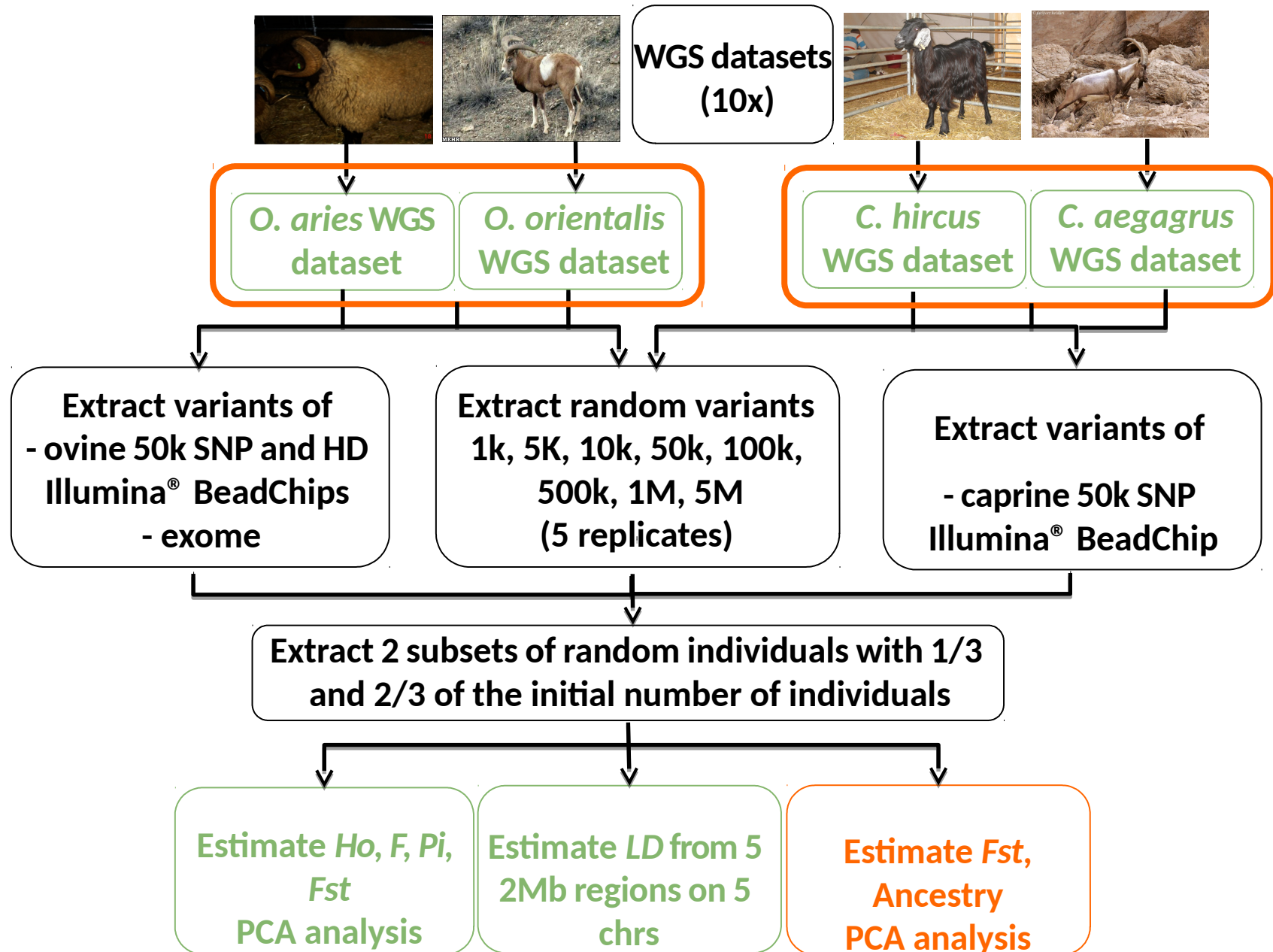
*C. aegagrus*

WGS dataset n=18

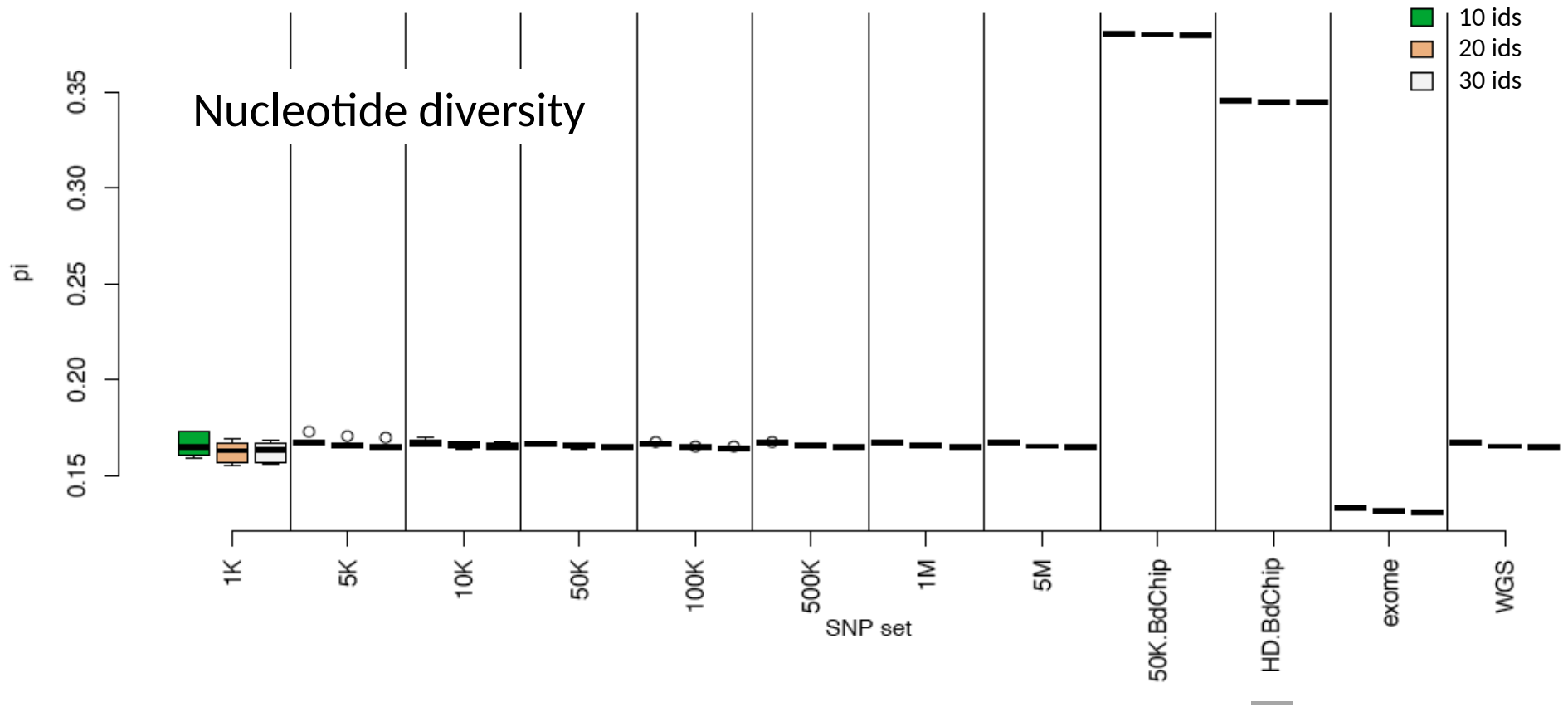
Divergence 10 Kya

Divergence 10 Kya

# Building datasets for assessing global diversity



# Estimating global neutral diversity

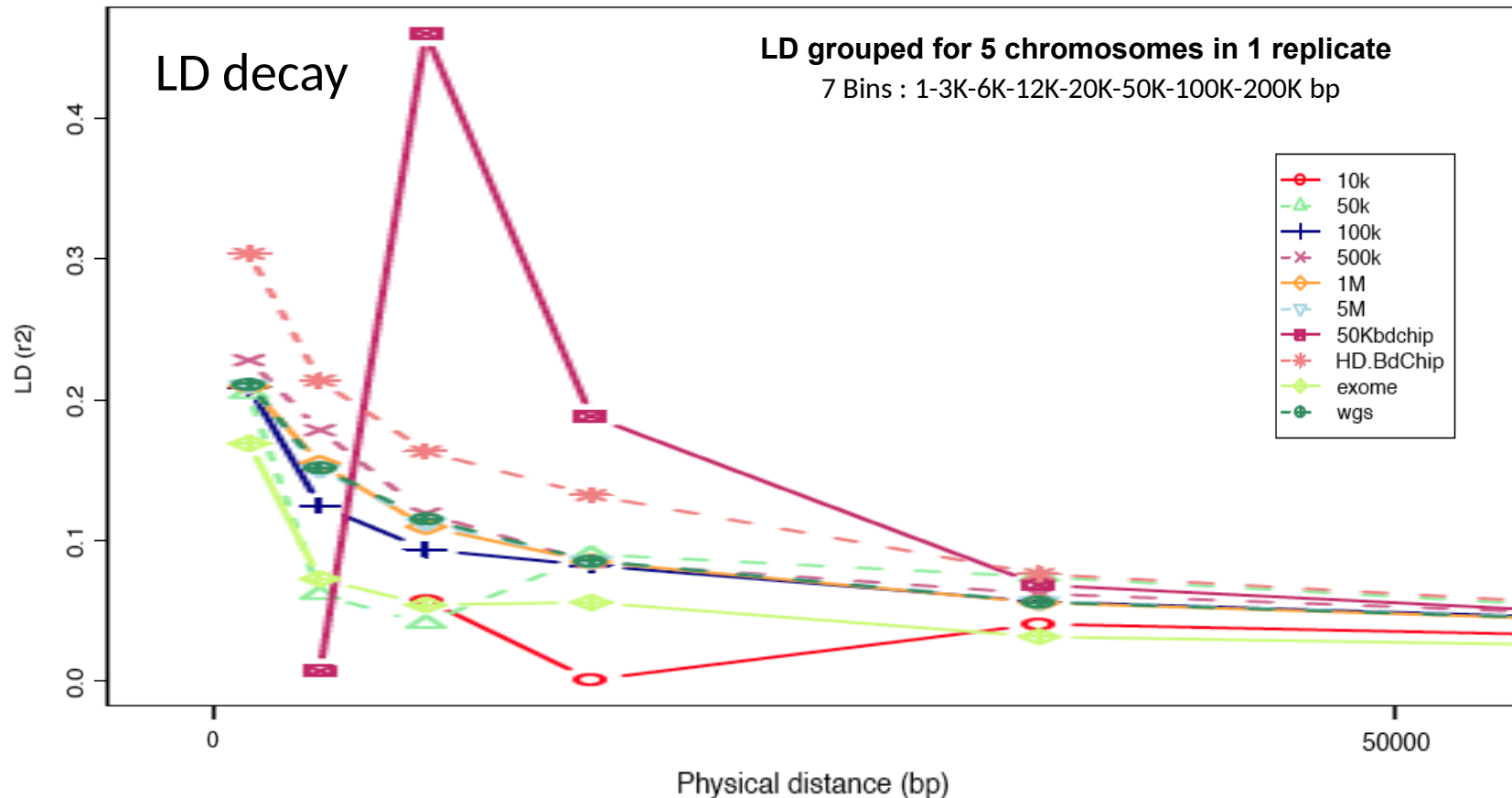


Random panels of 5K - 10K random SNPs for precise and unbiased estimates ( $\pi$ ,  $H_o$ ,...)

Ascertainment bias for chips & exome



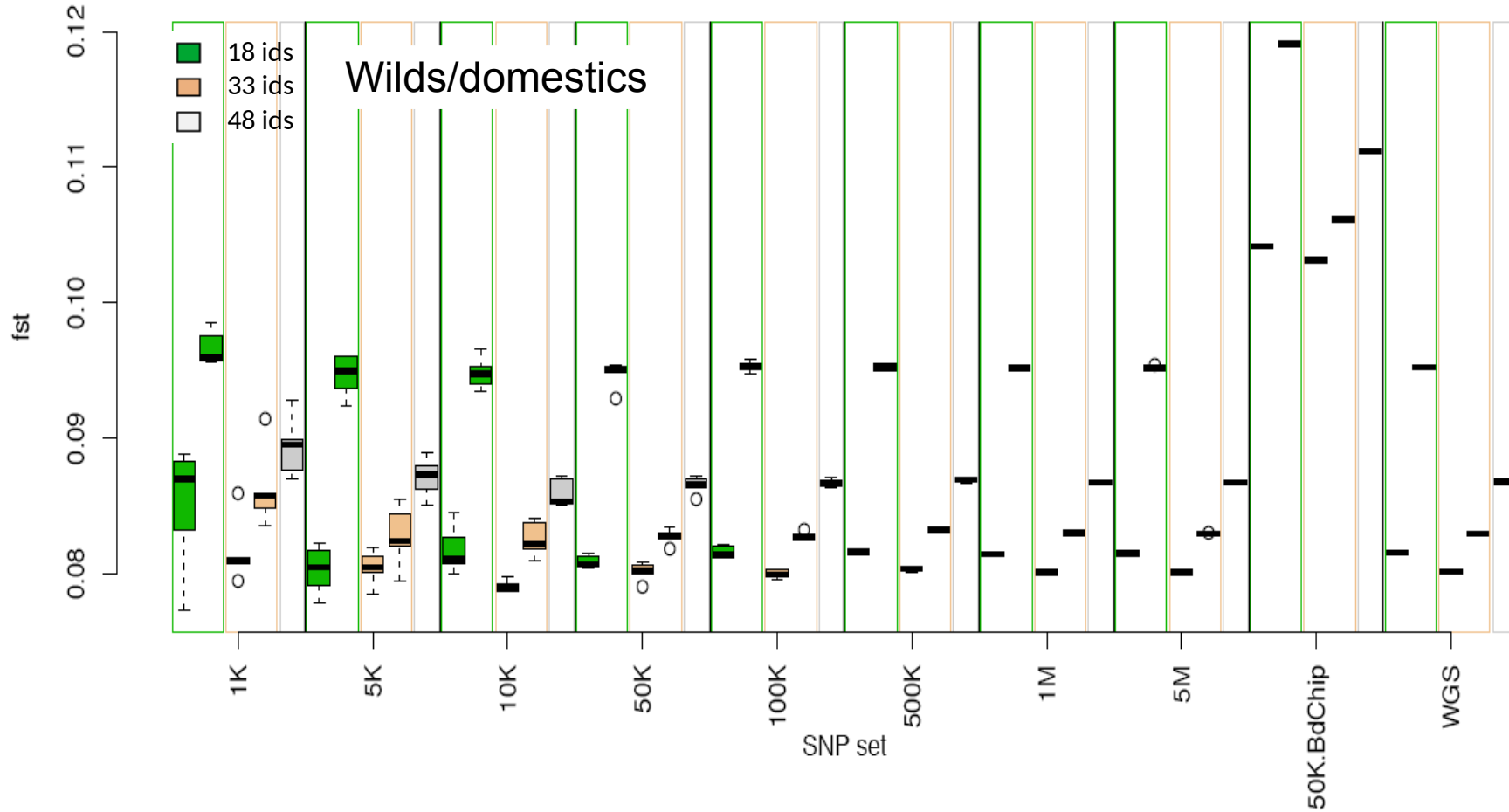
# Estimating Linkage Disequilibrium



Reliable LD estimates require a high SNP density (500K in *Capra* & 1M in *Ovis*)

BeadChips result in biased estimates

# Estimating Fixation index



Random panels of 50K – 100K SNPs for estimating  $F_{st}$

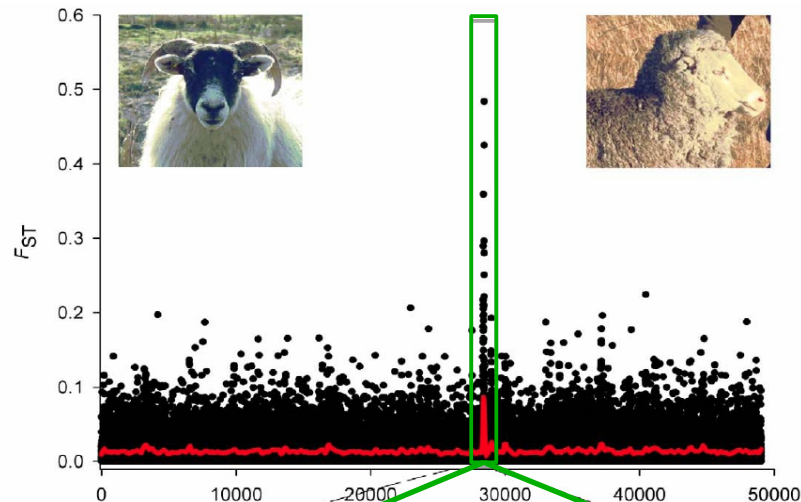
Ascertainment bias with BeadChips

Individual Sampling effect



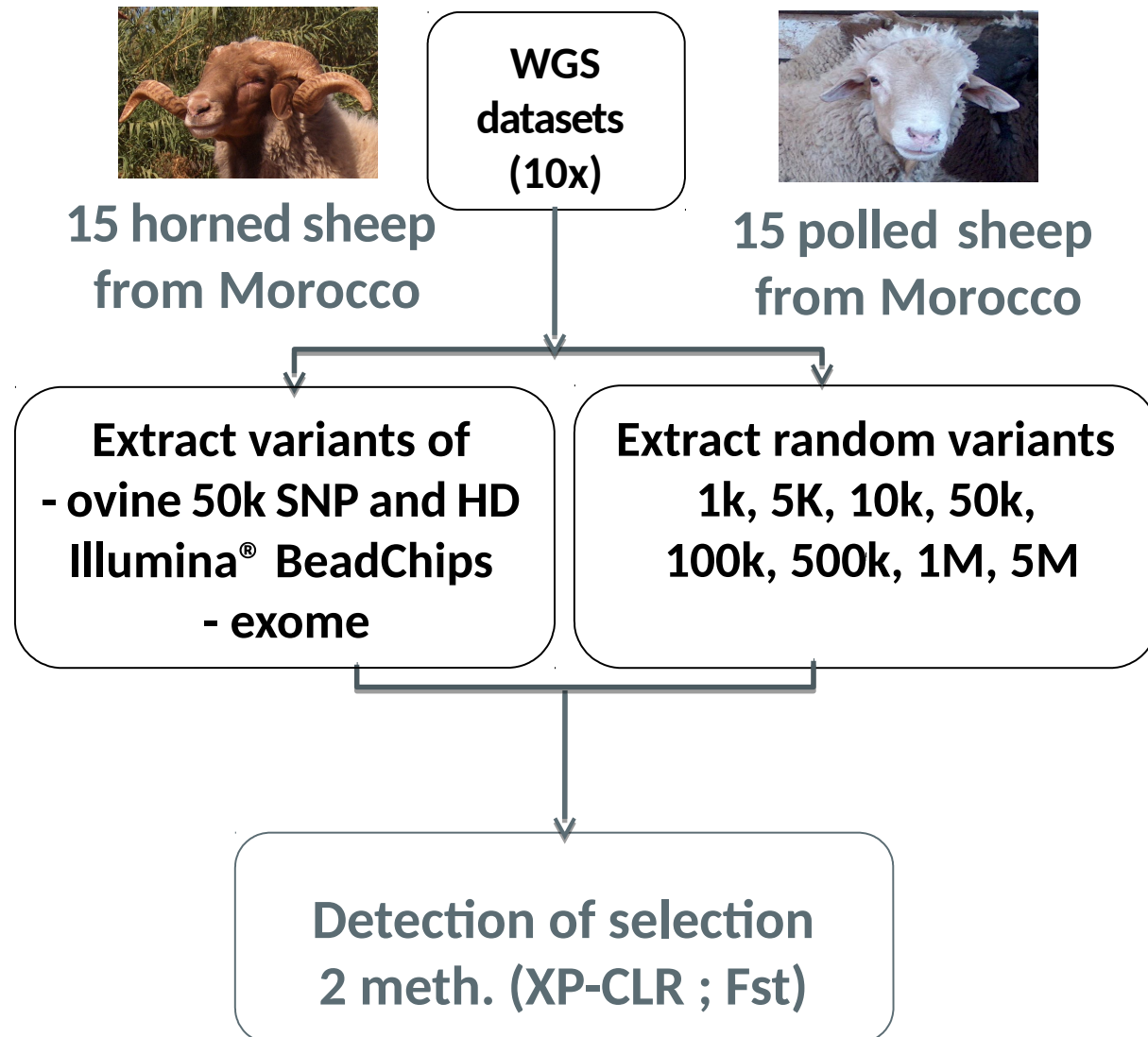
## Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection

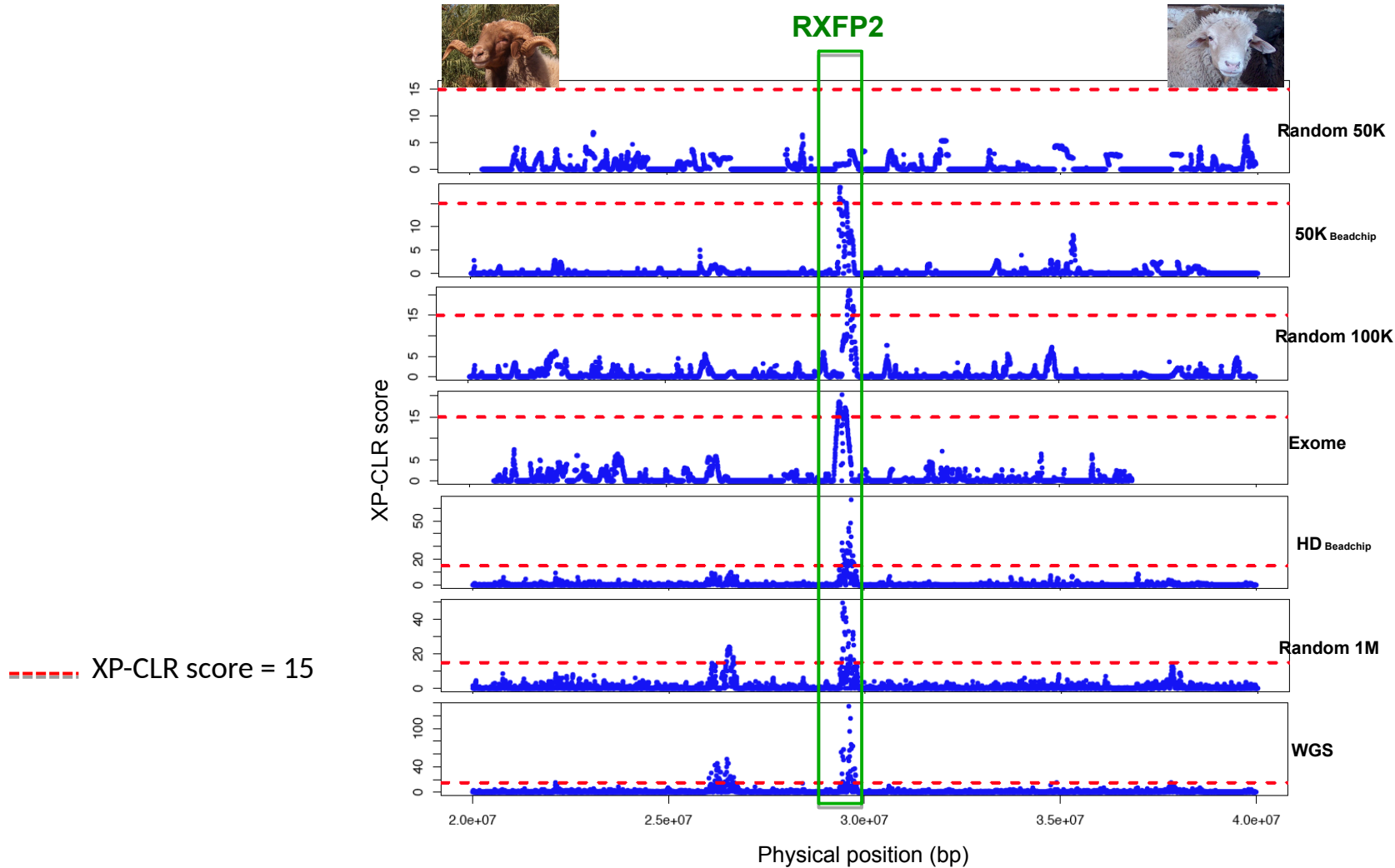
James W. Kijas<sup>1\*</sup>, Johannes A. Lenstra<sup>2</sup>, Ben Hayes<sup>3</sup>, Simon Boitard<sup>4</sup>, Laercio R. Porto Neto<sup>1</sup>, Magali San Cristobal<sup>4</sup>, Bertrand Servin<sup>4</sup>, Russell McCulloch<sup>1</sup>, Vicki Whan<sup>1</sup>, Kimberly Gietzen<sup>5</sup>, Samuel Paiva<sup>6</sup>, William Barendse<sup>1</sup>, Elena Ciani<sup>7</sup>, Herman Raadsma<sup>8</sup>, John McEwan<sup>9</sup>, Brian Dalrymple<sup>1</sup>, other members of the International Sheep Genomics Consortium<sup>10†</sup>



Chr 10: 20M - 40Mbp

# Building datasets for detecting adaptive genes





Detection of RXFP2 signal requires random panels  $\geq 100\text{K}$  SNPs  
 RXFP2 detected also by beadchips and exome  
 Other signals of selection may require higher densities ( $\geq 1\text{M}$ )

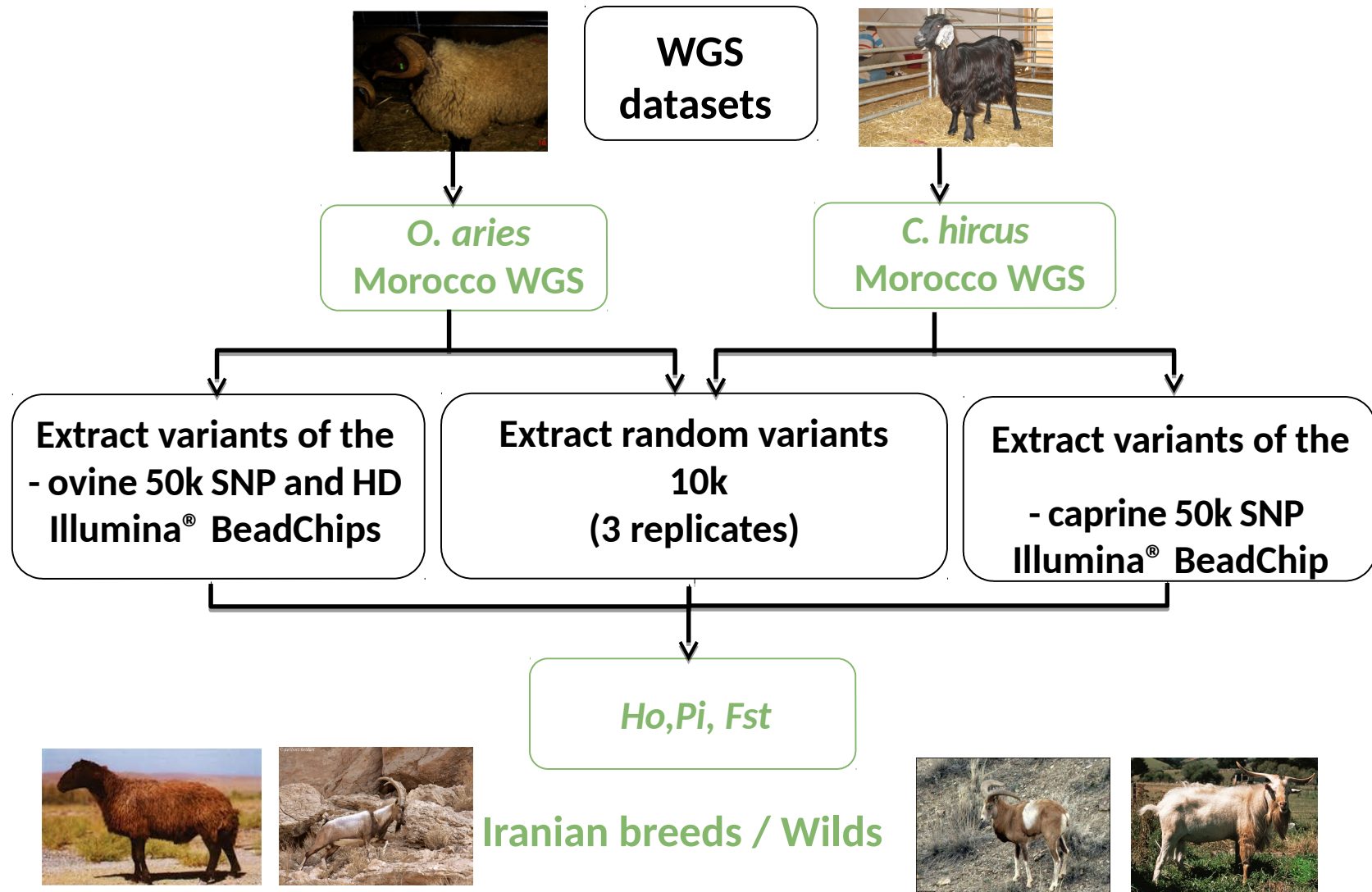


What about a generic ('universal') panel for studying genome diversity ?

Ascertainment bias of SNPs panels defined on industrial breeds

Ability of a random panel to reliably estimate genome diversity in any (many) breed/population ?

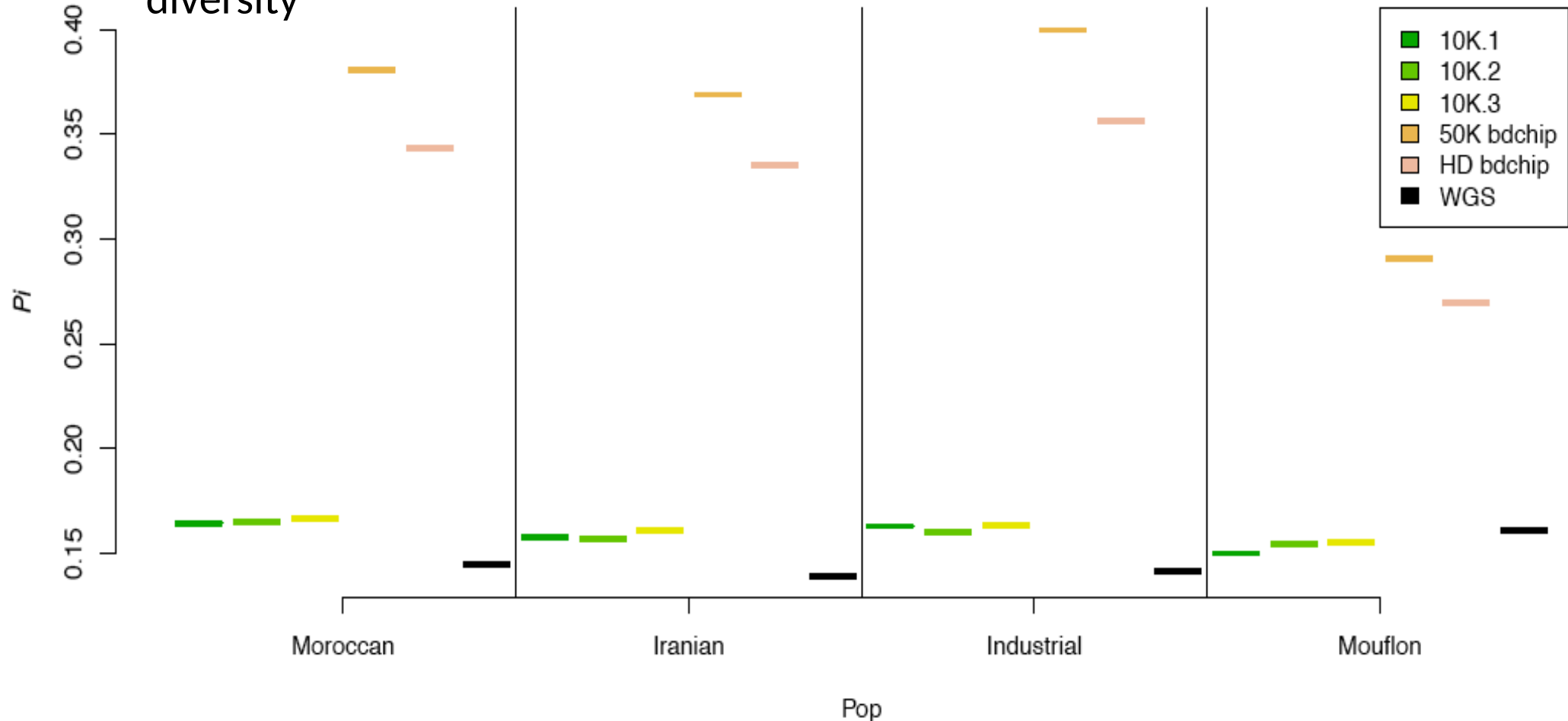
# Building datasets for assessing 'universality' of random panels



# 'universality' of 10k Moroccan random panels



nucleotide  
diversity



random panels are less biased and more precise than 50K-chips

## 2- Estimating genomic resources

Global diversity of wild relatives, local & industrial breeds

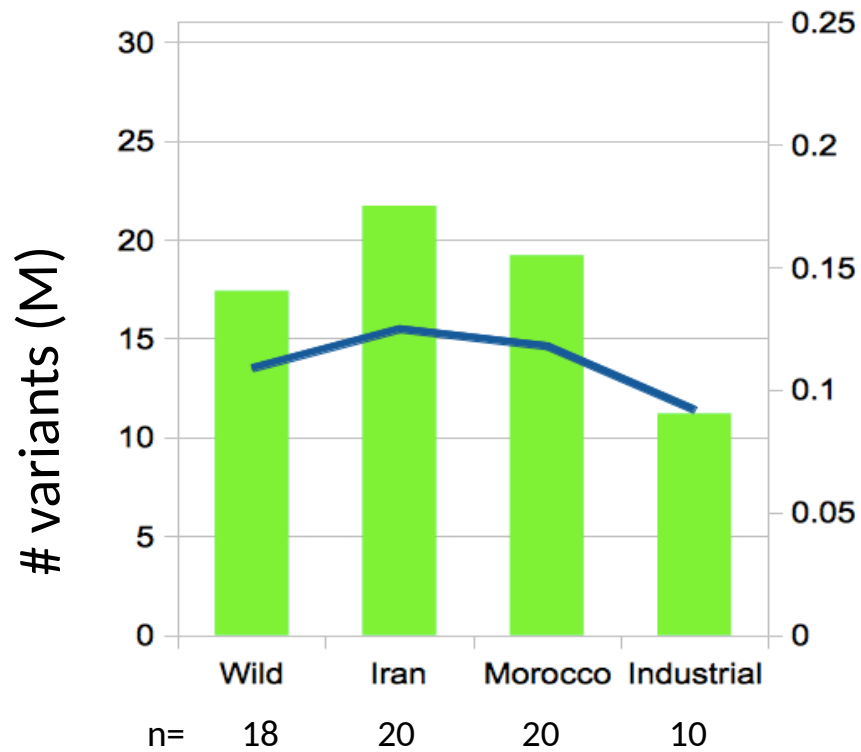
Levels of inbreeding

Where are the reservoirs of alleles ?

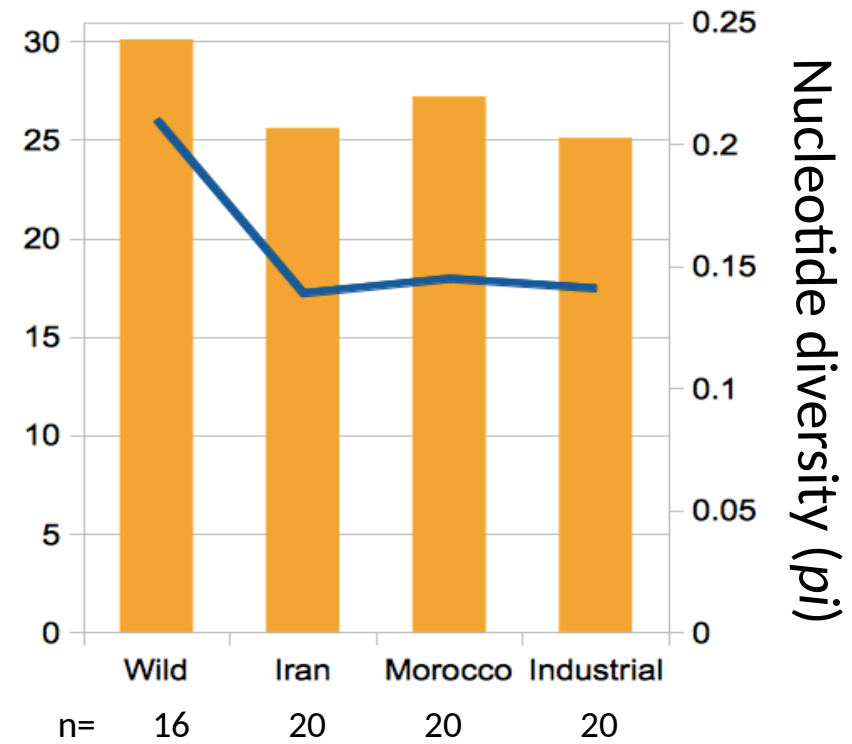
Value of indigenous breeds & wild relatives ?

# Genomic diversity

Goat

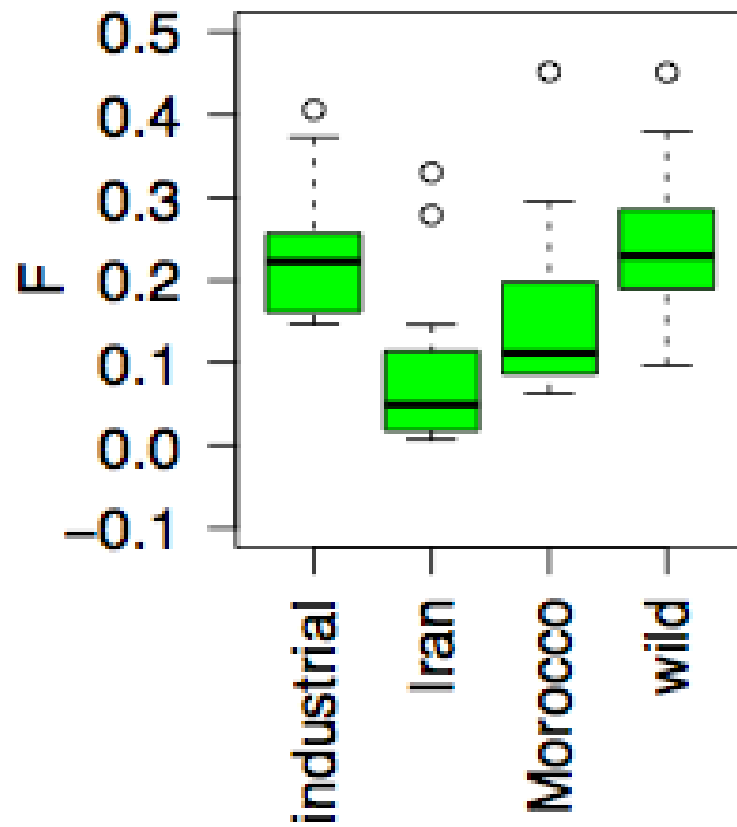


Sheep

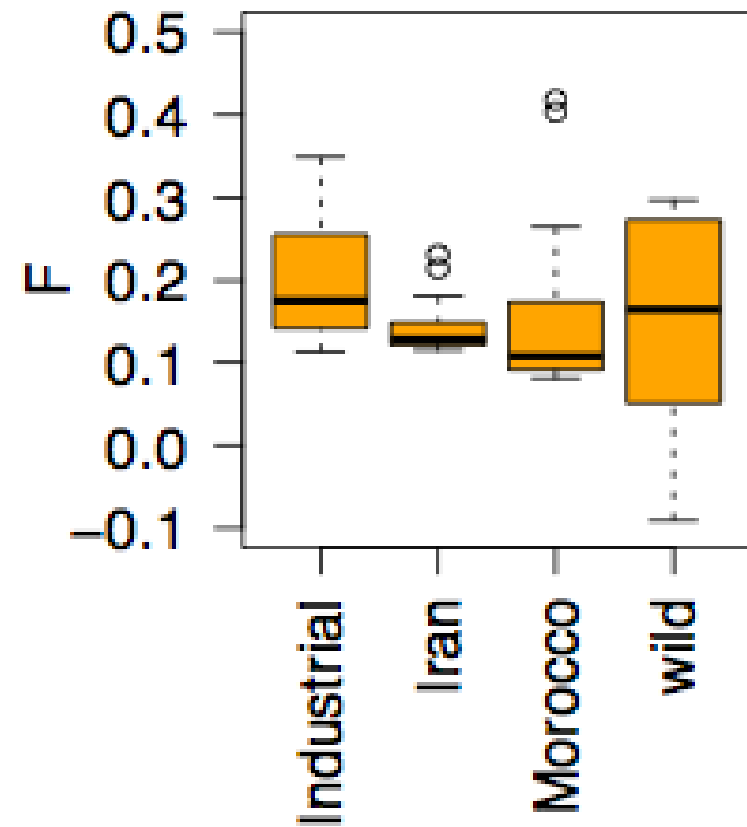


# Levels of inbreeding

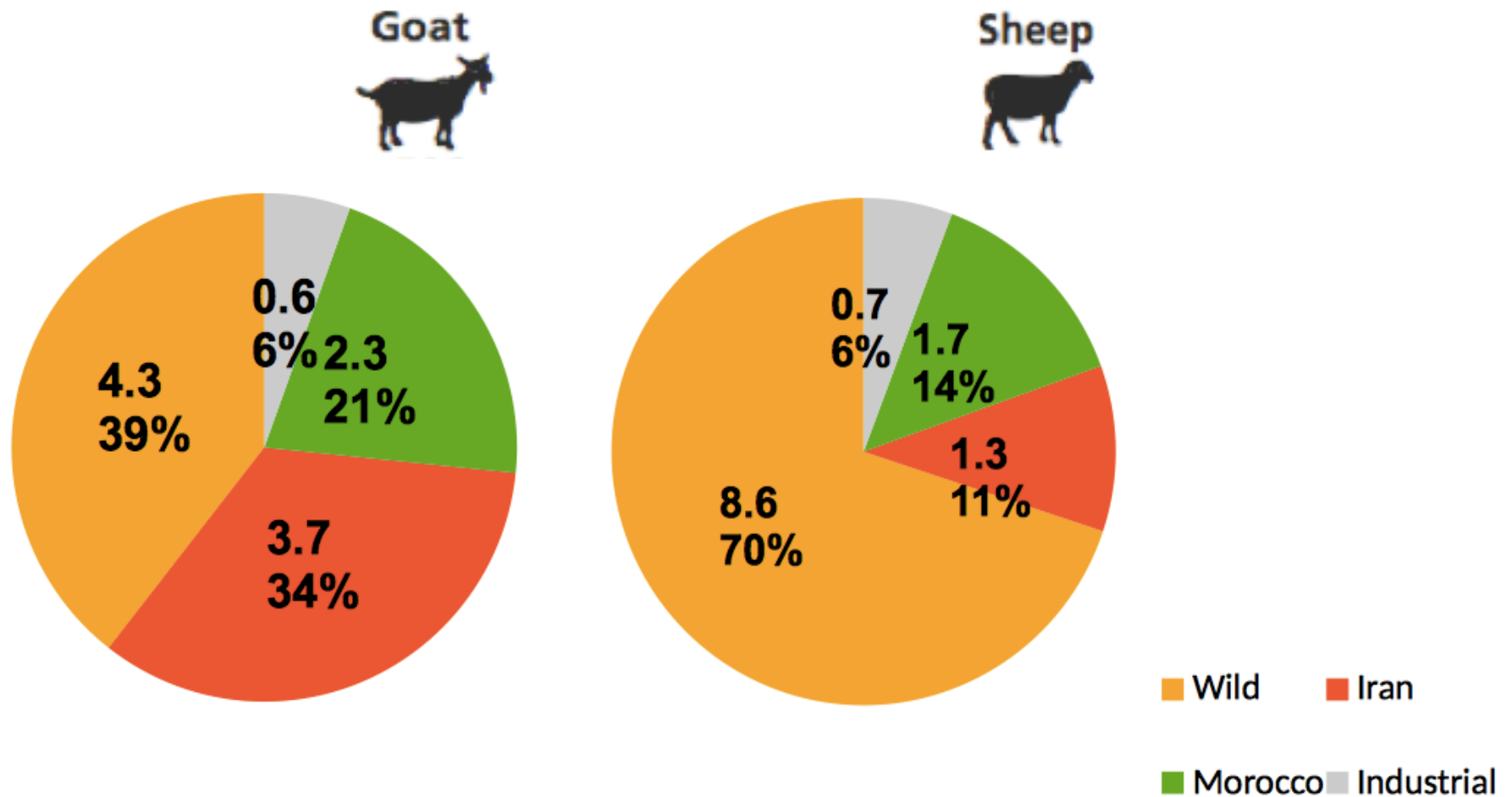
Goat



Sheep



# Where are the reservoirs of alleles ?



# of private alleles (M SNPs)

### **3- Local adaptation and genomic resources**

Identifying adaptive genes

combining neutral & adaptive diversity for defining conservation priorities



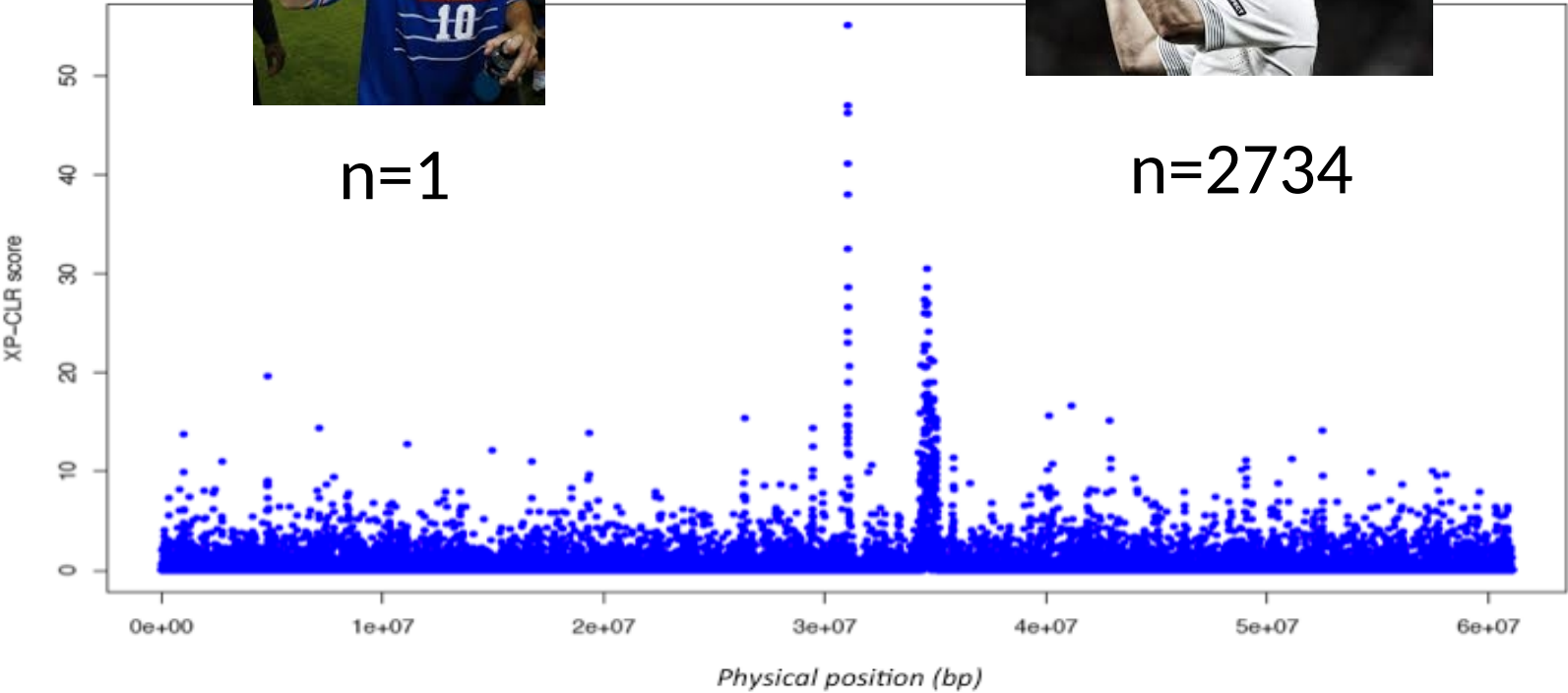
# Selection signatures for competitive ability



n=1



n=2734



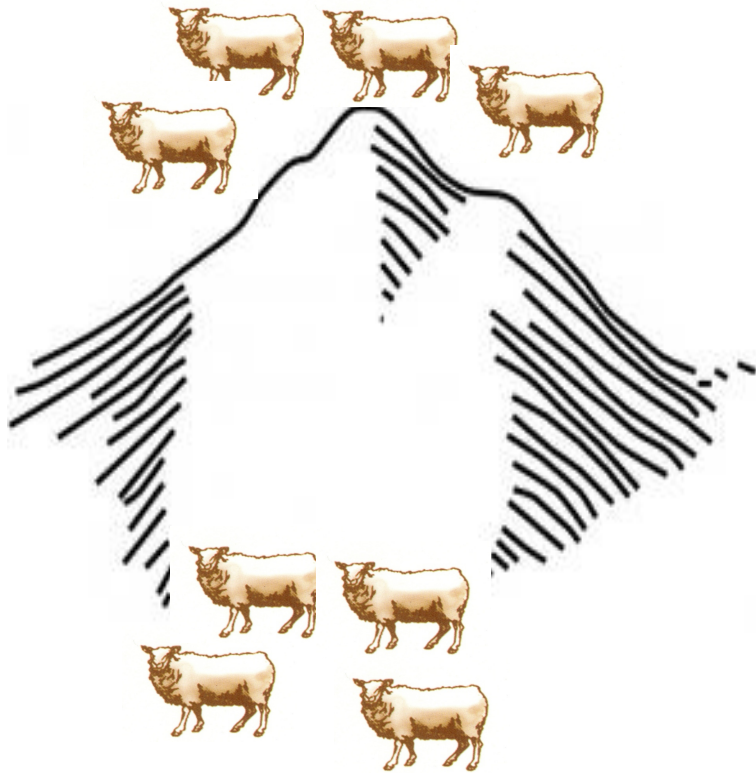
Project funding



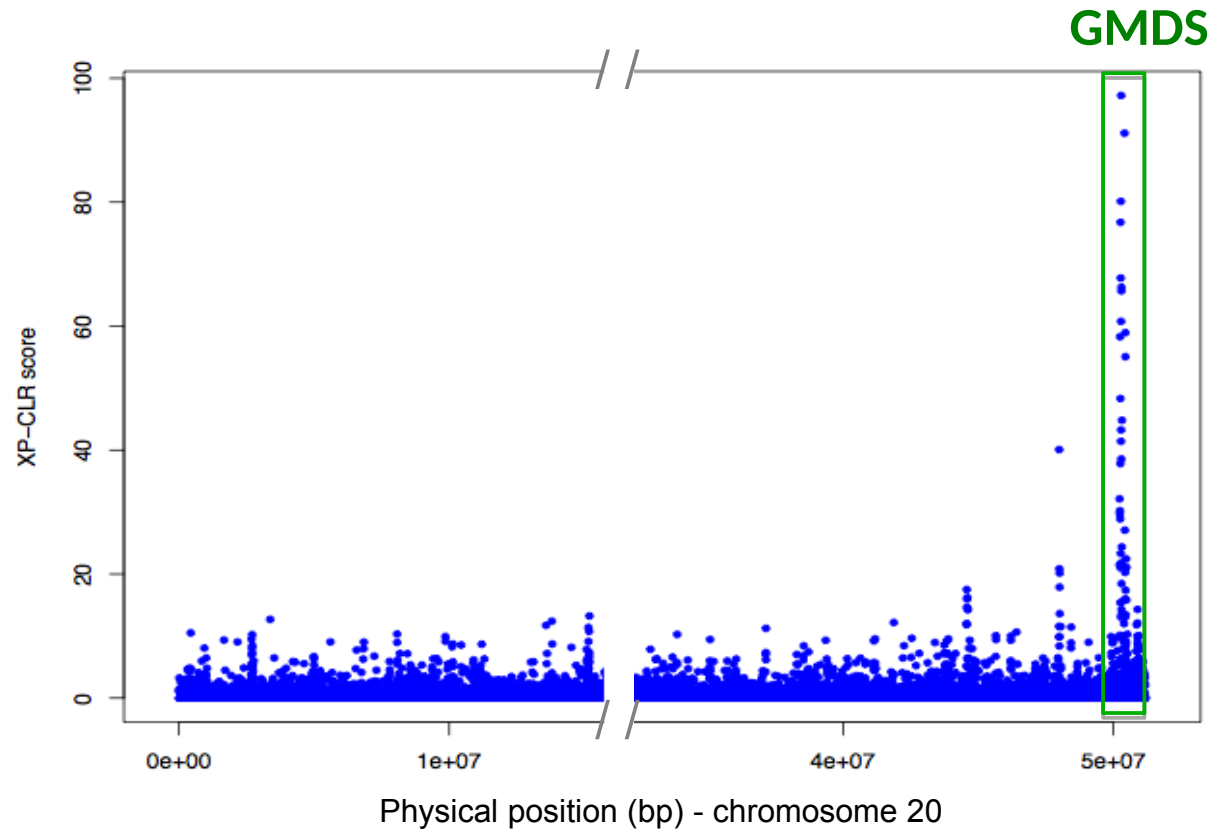
# Selection signatures related to altitude



>1433 m , n=25



< 219 m , n=25



*GDP-D-mannose 4,6-dehydratase* participates in fructose and mannose metabolism

# Conservation priorities: combining neutral & adaptive diversity

- Strategies at the population level:

Populations with the highest diversities

Populations with the highest pairwise diversity  
(complementarity principle)

- Neutral / Adaptive diversity

- Accounting for the adaptive value of breeds:

Population Adaptive Index (PAI) : % of 'adaptive loci' with allelic freq  $\neq$  from those in other populations

# Conservation priorities: combining neutral & adaptive diversity

Strategies:

N: Populations with the highest neutral diversities

NC: Populations with the highest pairwise neutral diversity

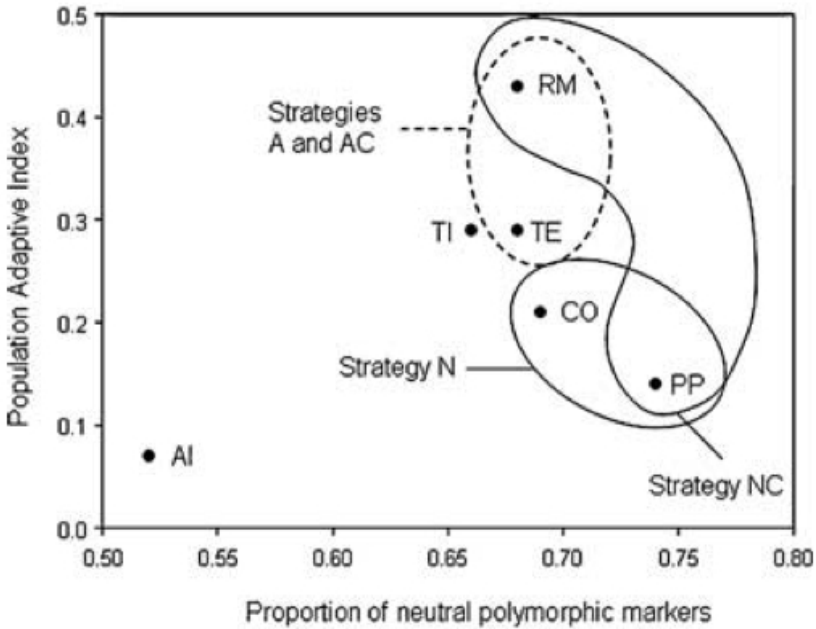
A: Populations with the highest adaptive diversity

AC: Populations with the highest pairwise adaptive diversity

# Conservation priorities: combining neutral & adaptive diversity

<i>Population<sup>a</sup></i>	<i>Proportion of polymorphic loci<sup>b</sup></i>	<i>Nei's gene diversity<sup>b</sup></i>	<i>PAI<sup>c</sup></i>
<b>Common frog</b>			
AI	0.52	0.18	0.07
CO	0.69	0.23	0.21
PP	0.74	0.25	0.14
RM	0.68	0.23	0.43
TE	0.68	0.23	0.29
TI	0.66	0.22	0.29

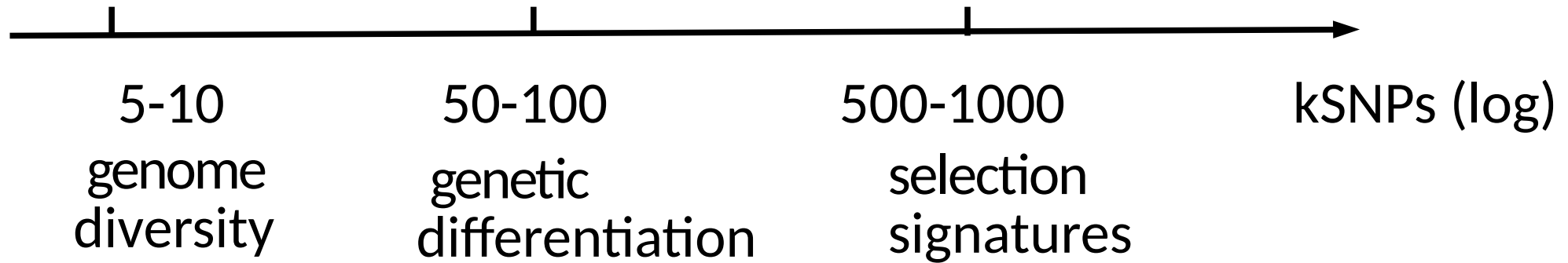
	<i>Common frog</i>					
	<i>AI</i>	<i>CO</i>	<i>PP</i>	<i>RM</i>	<i>TE</i>	<i>TI</i>
AI	-	0.81	0.90	0.83	0.83	0.76
CO	0.21	-	0.95	0.87	0.90	0.82
PP	0.21	0.36	-	0.95	0.93	0.93
RM	0.50	0.50	0.57	-	0.91	0.89
TE	0.36	0.50	0.42	0.71	-	0.89
TI	0.21	0.29	0.42	0.64	0.50	-



Bonin et al. 2007 Conserv. Biol.

# Conclusions

- Random SNPs panels are efficient



- Taking into account 'adaptive' diversity in conservation priorities:  
From outliers loci to adaptive pathways:  
global adaptive potential / targeted adaptive traits ?

- Wild relatives and indigenous breeds do constitute a reservoir of genomic resources.  
How setting up whole genome approaches for characterization and prioritization ?

# Acknowledgements

## The Next Generation Scientists



**Florian Alberto**    **Badr Benjelloun**

**Paolo Ajmone-Marsan**  
**Adriana Alberti**  
**Filippo Biscarini**  
**Aurélie Bonin**  
**Fred Boyer**  
**Mike Bruford**  
**Laura Clarke**  
**Licia Coli**  
**Stefan Engelen**  
**Paul Flicek**  
**Stéphane Joost**  
**James Kijas**  
**Raffaele Mazza**  
**Saeid Naderi**  
**Riccardo Negrini**  
**Pablo Orozco-terWengel**  
**Hamid Rezaei**  
**Alessandra Stella**  
**Ian Streeter**  
**Sylvie Stucky**  
**Pierre Taberlet**  
**Wahid Zamani**



**NEXTGEN consortium**