High throughput SNP discovery in the pig using the Illumina Genome Analyzer and characterization of the porcine HapMap panel using the Illumina Porcine 60K iSelect TM Beadchip

Martien Groenen

ANIMAL SCIENCES GROUP
WAGENINGEN UR

Illumina Agriculture Seminar Series
August 3, 2009, Madison WI

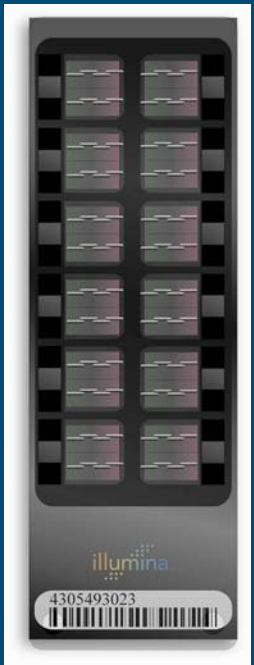
Animal Breeding &
Genomics Centre

Overview

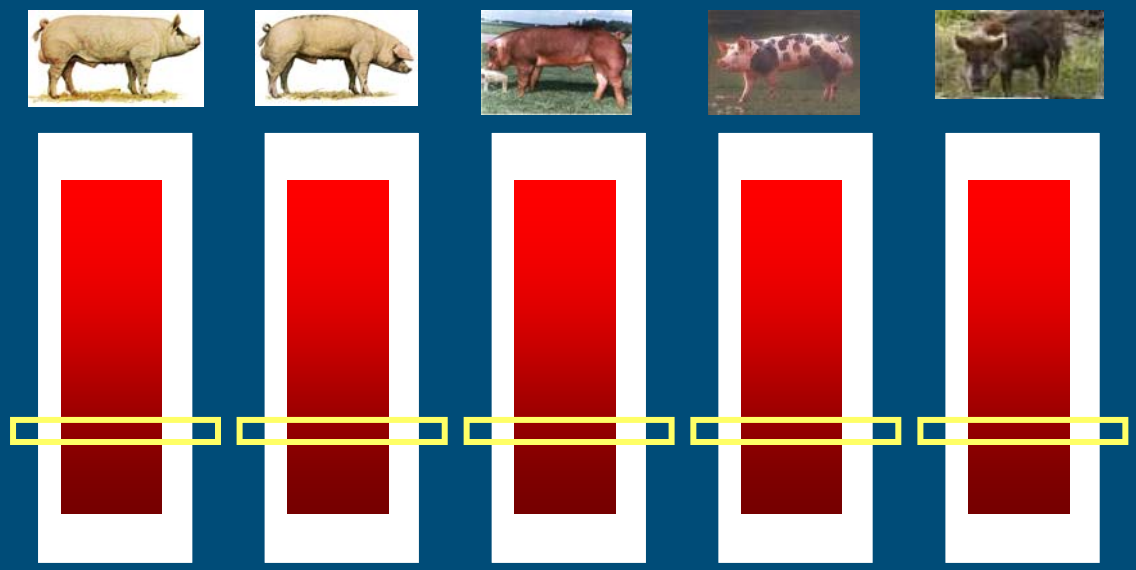
- SNP discovery and chip design
- Statistics of the 60 K chip
- Hapmap data: MDS and clustering
- Hapmap data: Haploype diversity, haplotype sharing and selective sweeps
- Fom SNP genotyping towards resequencing

History

- Objective
  - Design of a 60K Illumina BeadChip
- Challenge
  - Limited number of SNPs available
  - Only half of porcine genome sequence available
  - Job had to be done between April/August 2008
- Approach
  - SNP identification by sequencing reduced representation libraries (RRL) on an Illumina Genome Analyzer



Overview sequencing approach



- 5 Reduced representation libraries (RRL)
- Restriction digest of DNA
- Separate on acrylamide gels
- Isolate 150-200 bp fragments

→ Rough estimate of MAF within breeds

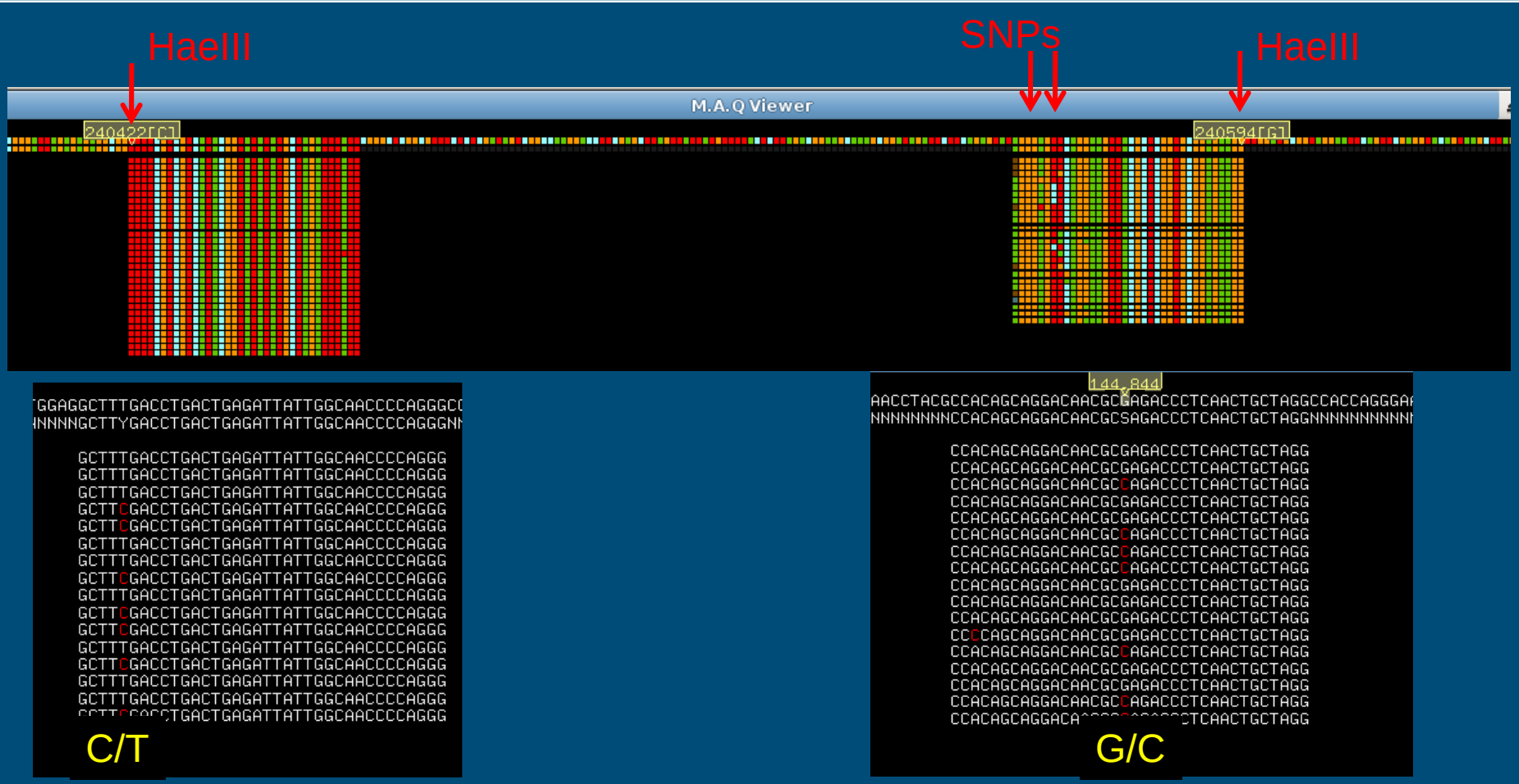
30x → Highly reliable SNP identification

RRLs results

RRL	Genome %	Total No Reads (million)	No Reads after filtering (million)
HaeIII 160-200	1	67.1	57.8
AluI 100-150	2	88.6	83.4
AluI 150-200	3	145.9	95.8
MspI 100-200	1	69.1	52.3
DraI 180-260	2	100.3	67.7
TOTAL	**9**	**481**	**357**

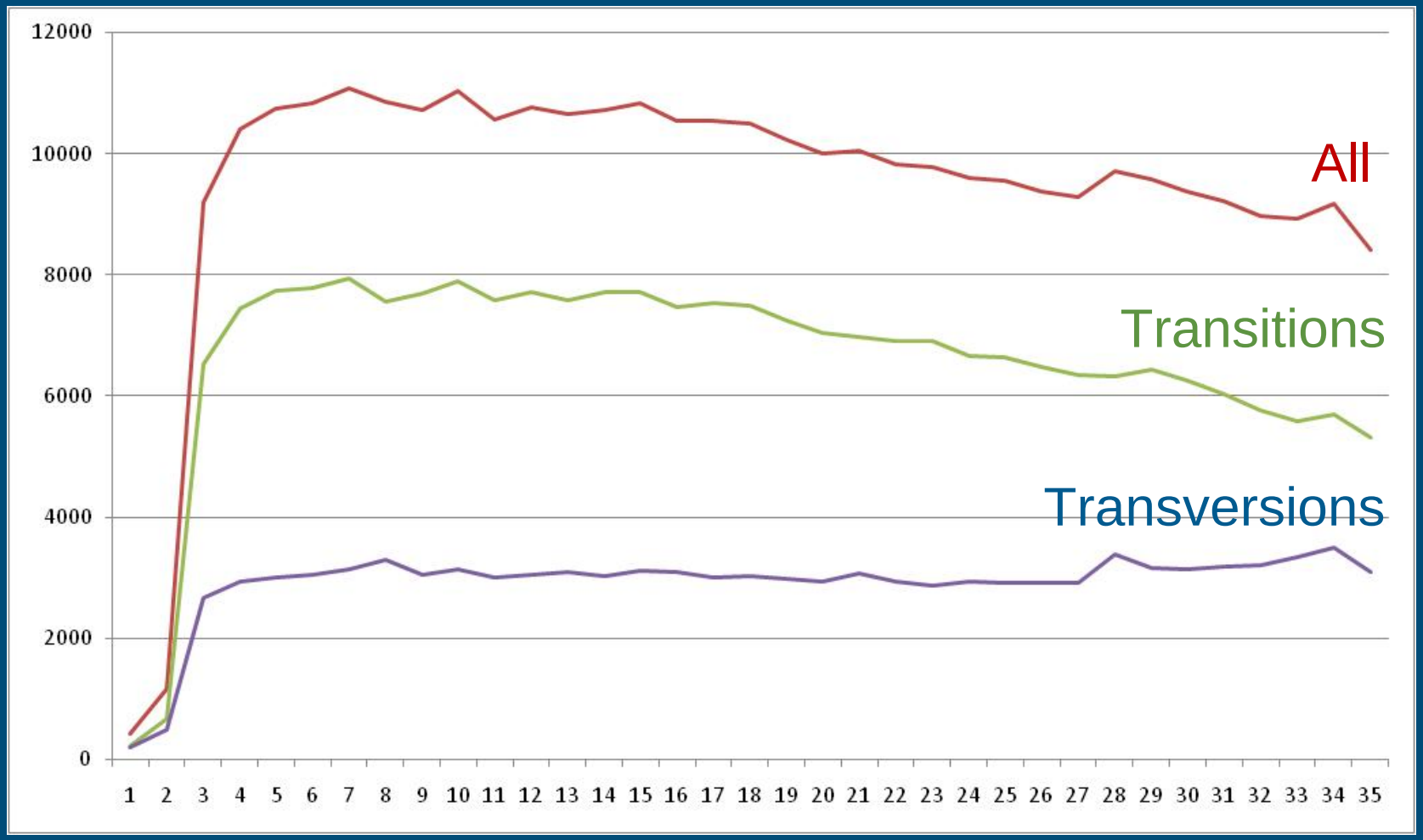
=12.8 billion bp

Align to reference genome and identify SNPs using MAQ



SNP called if minor allele seen at least 3x

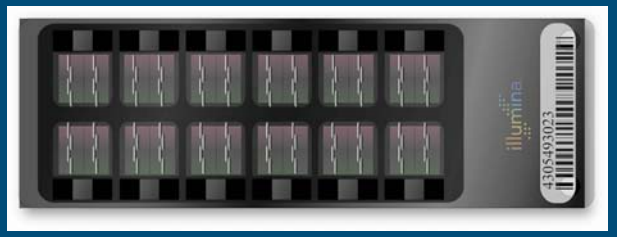
SNP distribution on Illumina GA reads



All porcine SNPs currently available

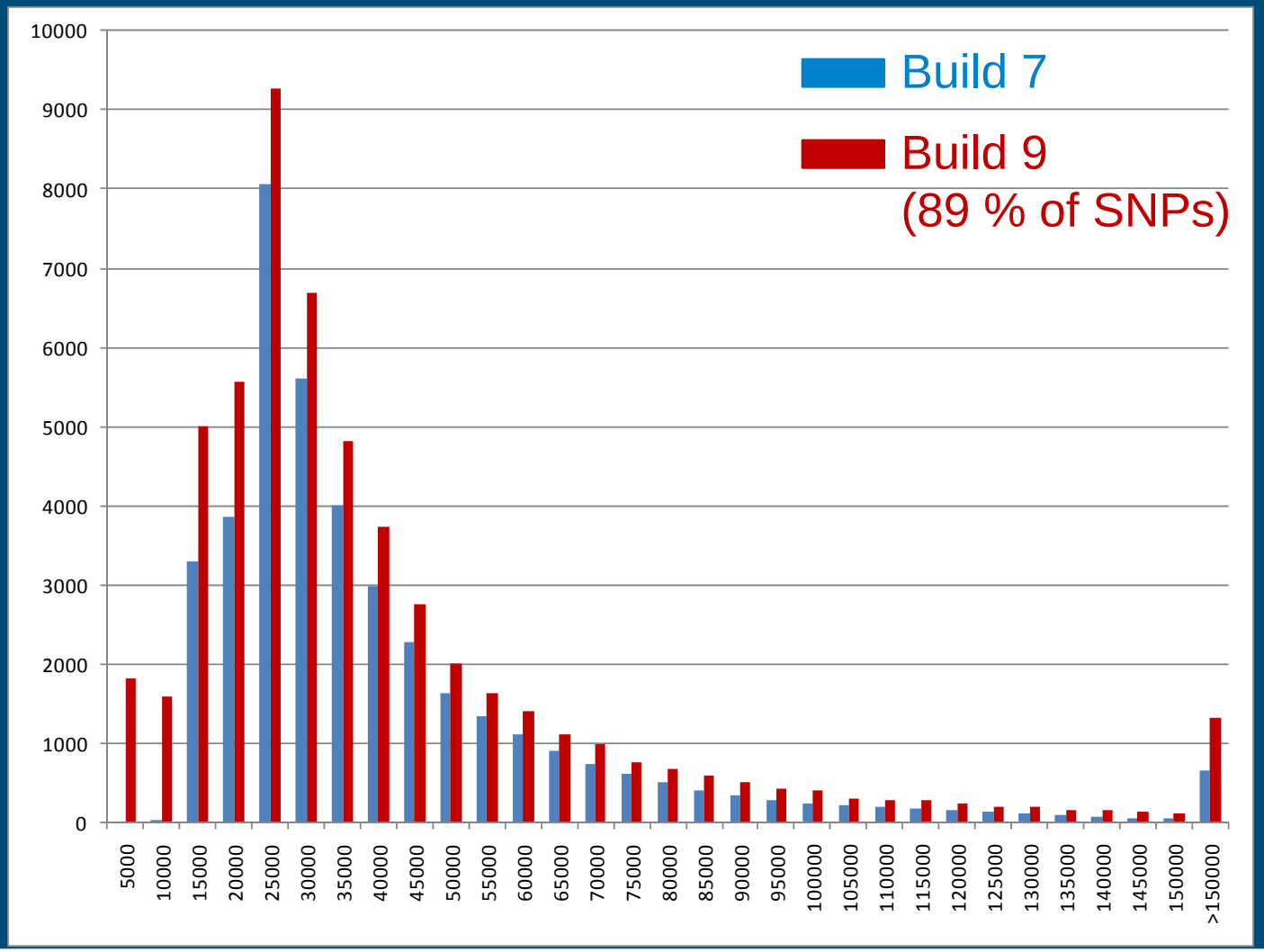
SNP source	Number
HQ Solexa SNPs WU*	333,000
454 sequencing MARC	110,000
7K iSelect chip AU-RI	5,500
dbSNP	17,700
INRA (Sanger seq)	61,700
Cambridge University	14,900
Total No SNPs	543,000
Total Unique SNPs	**510,000**

*Plus additional 58,000 SNPs where minor allele seen twice



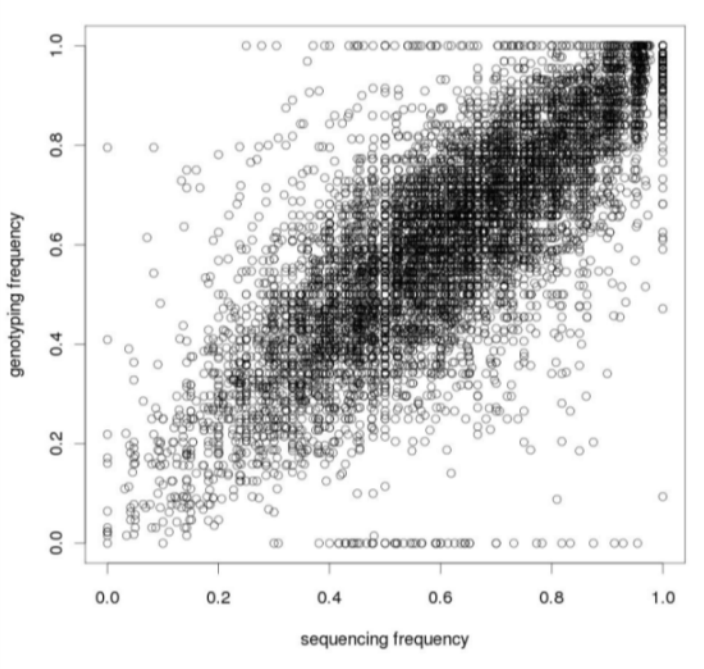
- TOTAL #SNPs submitted for the chip: 72,000
  - 70 % mapped on build 7
  - 7 % predicted
  - 23 % unmapped

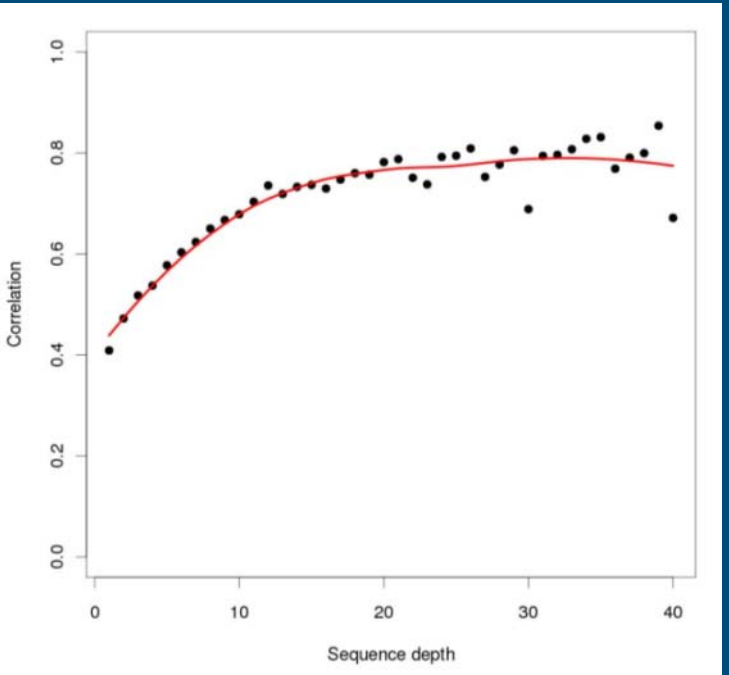
SNP distribution



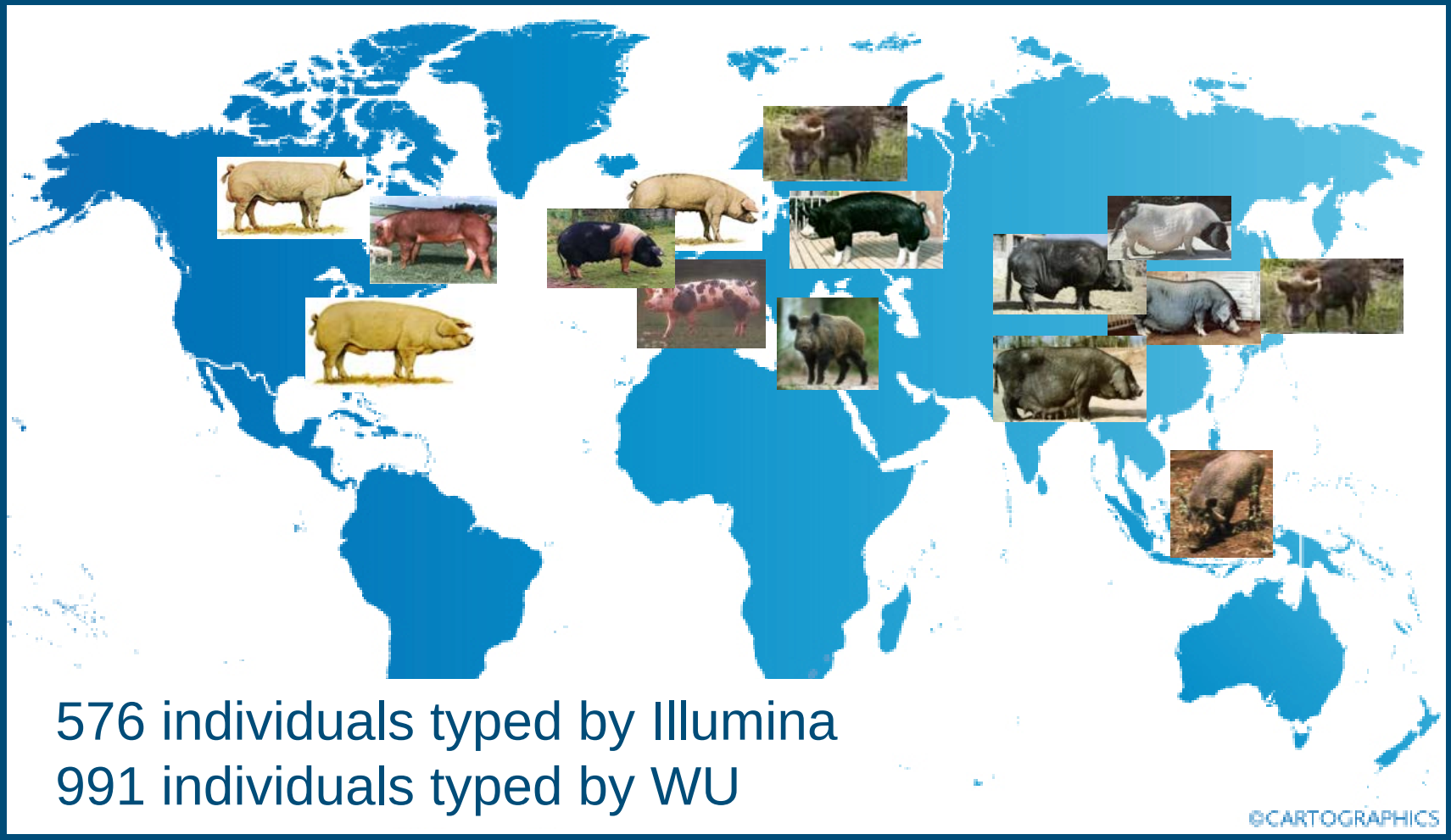
Porcine 60K BeadChip: Excellent performance

- Final total number of SNPs on the chip: 62,163
- Number of SNPs with MAF>0.05 in at least 1 breed: 59K
- Conversion rate overall: 94.8% (Illumina GA SNPs: 96 %)
- Good correlation between allele count based on GA sequencing and Geotyping





The porcine HapMap project



Phenotypic variation

Porcine Hapmap panel

Breed	Illumina	WU	Total
Landrace	79	33	112
Large White	129	28	157
Duroc	82	16	98
Pietrain	95	17	1112
Wild Boar	20	249	269
12 Chinese breeds	31	233	264
3 Synthetic	6	49	55
Hampshire	69	8	77
Berkshire	58	-	58
Museum samples	-	28	28
Other suiformes	6	41	47
Other european br.	-	289	289
Tobasco	1	-	1
TOTAL	**576**	**991**	**1567**

Including 165 animals of the discovery panel

Sus celebensis
Sus verrucosus
Sus barbatus
Pecari tajacu
Potomochoerus
Babyrousa
Sus cebifrons

Animal sequenced

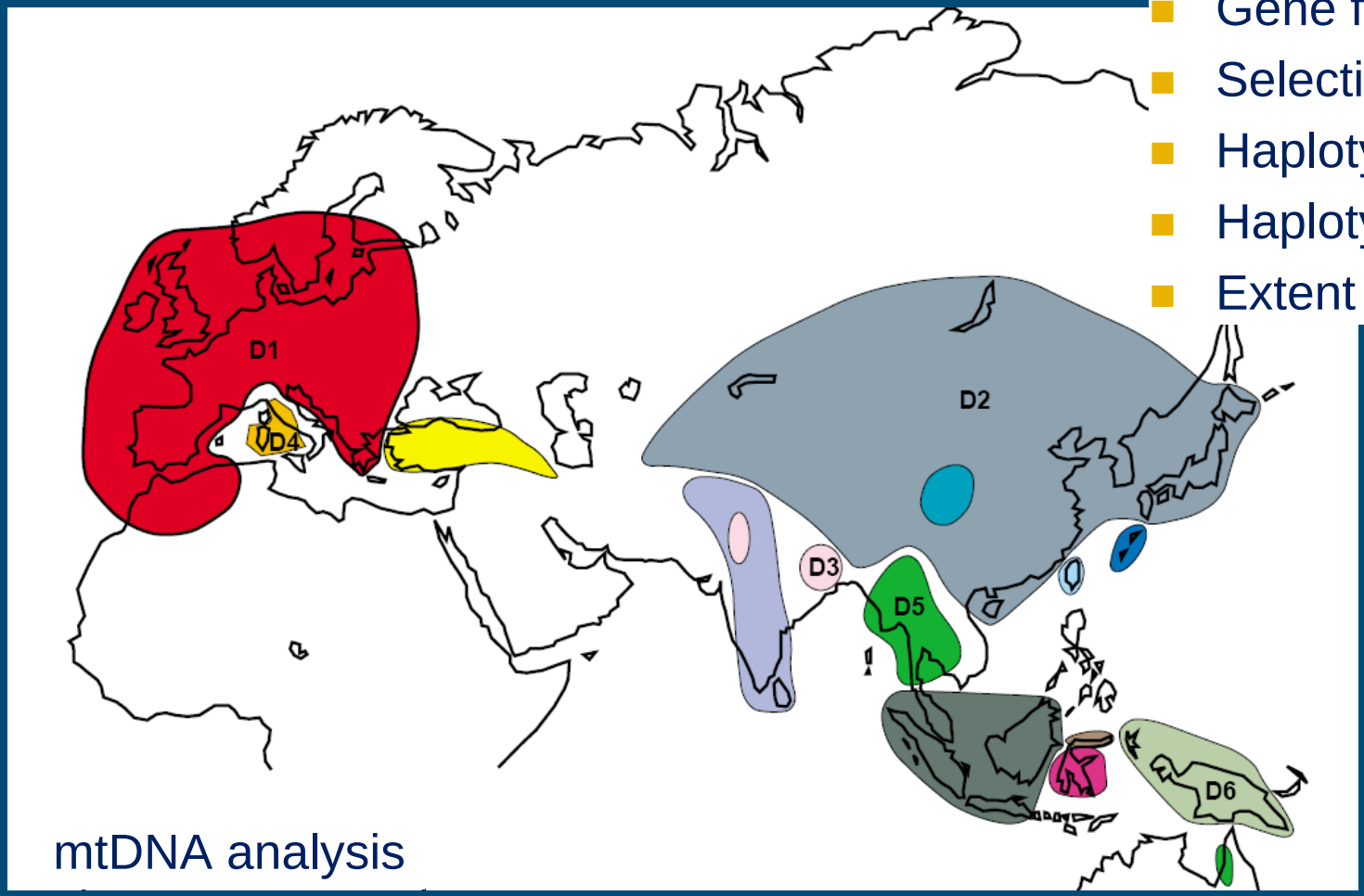
Use of the 60K chip: European vs Chinese breeds

- Biased towards common variants
- Biased towards SNPs in European breeds
- SNP density of the 60K chip
  - Sufficient for GWA and GWS in european breeds
  - Too low for Chinese breeds

Amaral et al. (2008) Genetics 179: 569

Domestication of the pig

- Gene flow
- Selective sweeps
- Haplotype diversity
- Haplotype sharing
- Extent of LD



mtDNA analysis

Greger Larson et al. (2005) Science 307:1618

Multi dimensional scaling and clustering

MDS plot Axis1-2: Division Europe-Asia (PLINK)

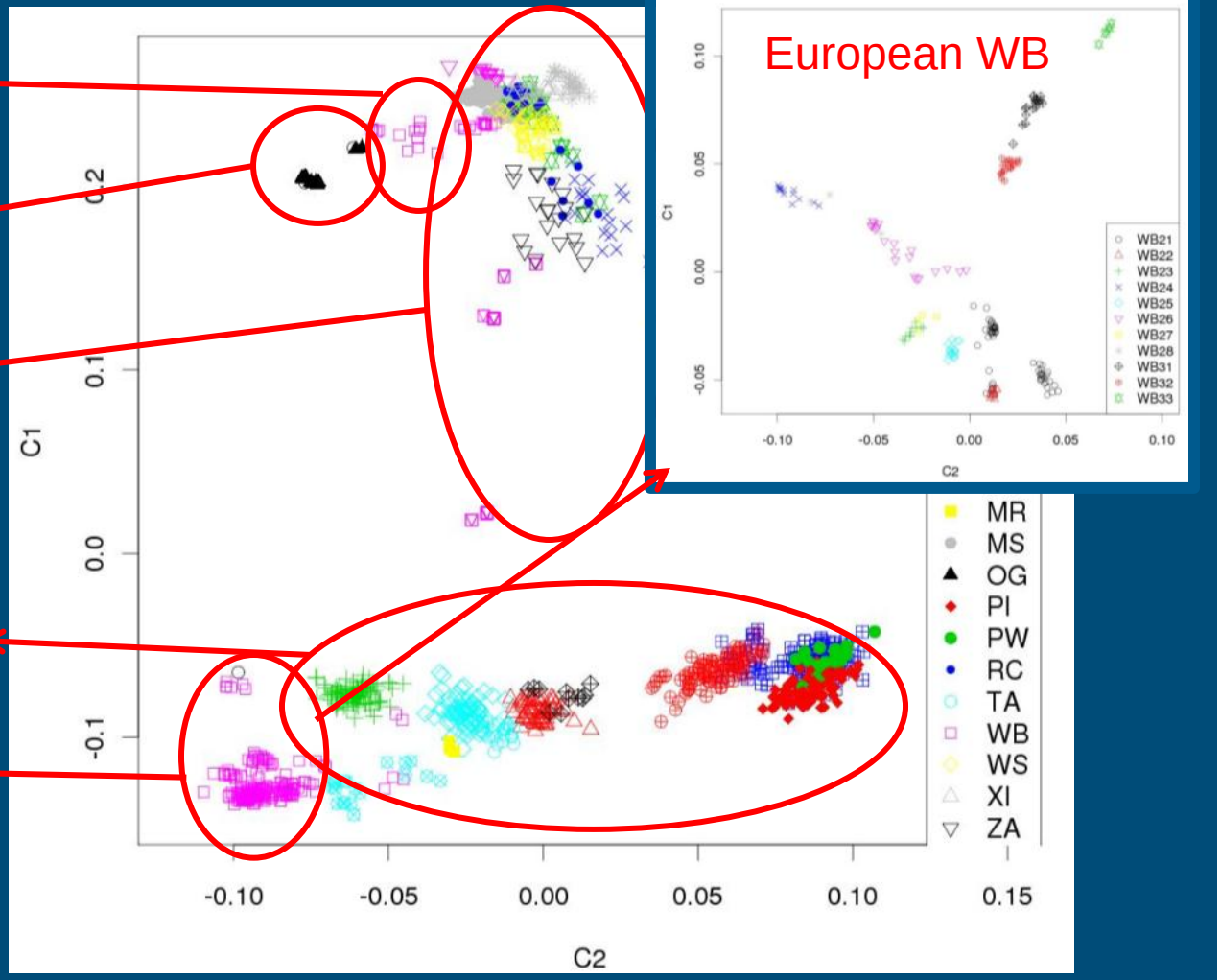
Asian WB

Outgroup
(other suidae)

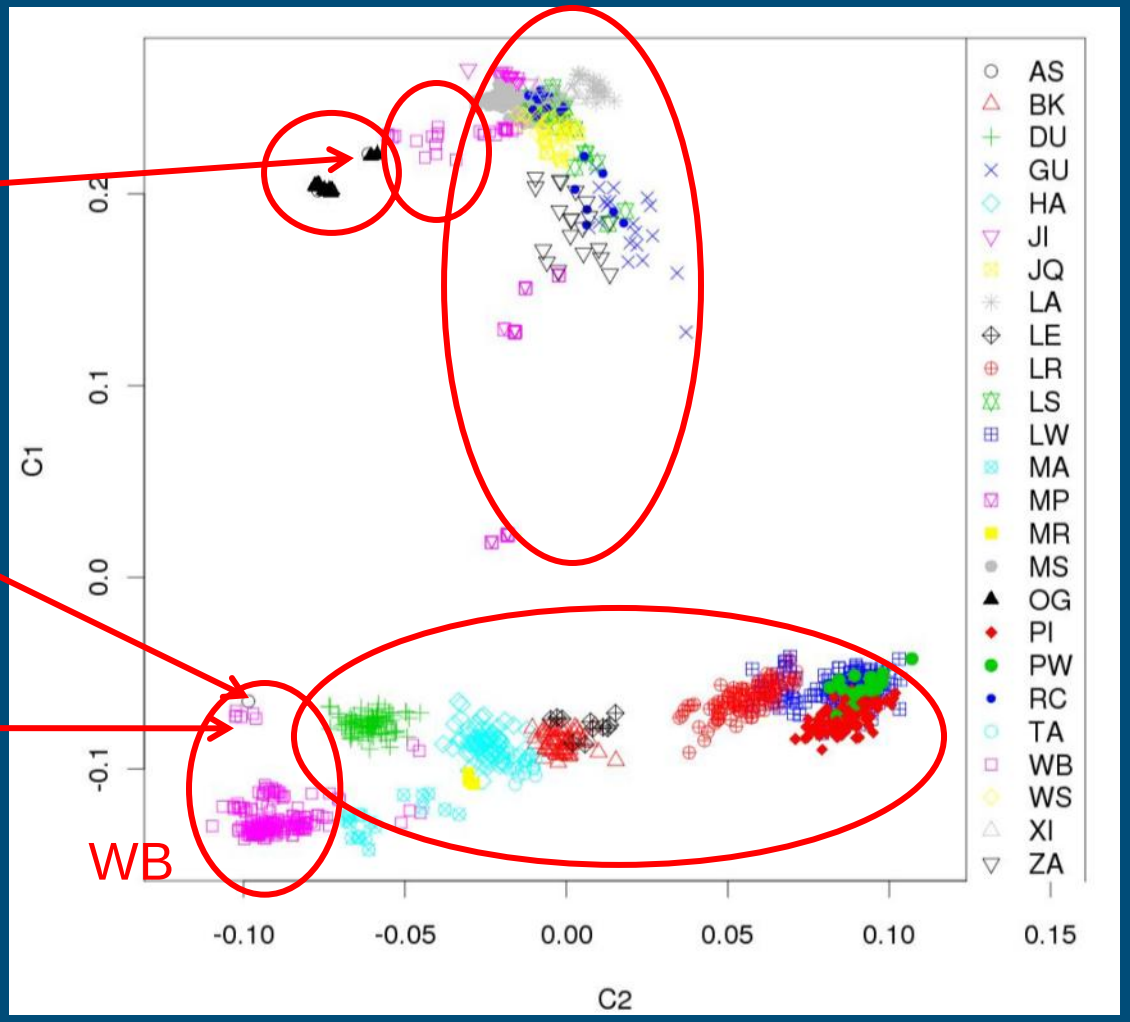
Asian breeds

European/US breeds

European WB



MDS plot Axis1-2: Division Europe-Asia (PLINK)



Outgroups
Other suidae

Other Suidae

Sus scrofa

Sus barbatus (11)

Sus cebrifons (1)

Sus celebensis (2)

Sus verrucosus (10)

~ 1 MY

~ 2 MY

Potomochoerus l. (1)

Potomochoerus p. (1)

~ 4 MY

Phacochoerus a. (2)

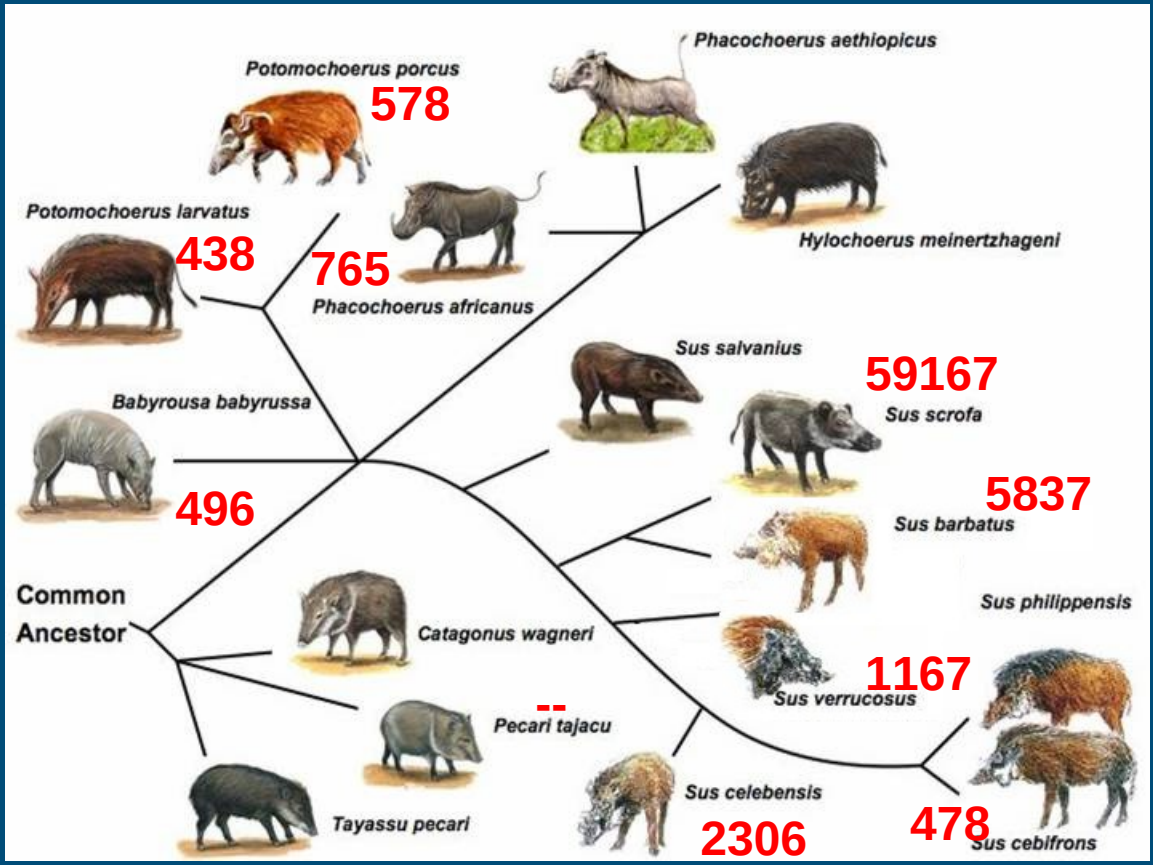
Babyrousa (1)

~ 10 MY

Pecari tajacu (1)

~ 70 MY

SNPs shared with other suidae



Species	Average heter.
Sus scrofa	23 %
Sus barbatus	3 %
Sus celebensis	2 %
Sus cebifrons	0.8 %
Sus verrucosus	0.6 %
Phacochoerus	0.8 %
Potomochoerus l.	0.7 %
Potomochoerus p.	0.9 %
Babyrousa b.	0.9 %

Ancestral allele identified for > 95% of the SNPs

Haplotype diversity, haplotype sharing and selective sweeps

Haplotype sharing between pig populations

Sharing is high between populations of the same breed, but not uniform across chromosomes

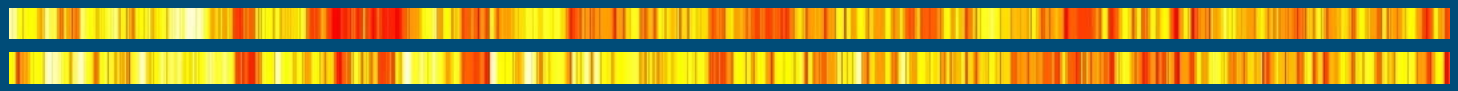
Pietrain/Pietrain
Duroc/Duroc

Sharing is low between Chinese and European breeds, but local exceptions exist

Duroc/Meishan

Sharing is intermediate between populations of several 'white' breeds

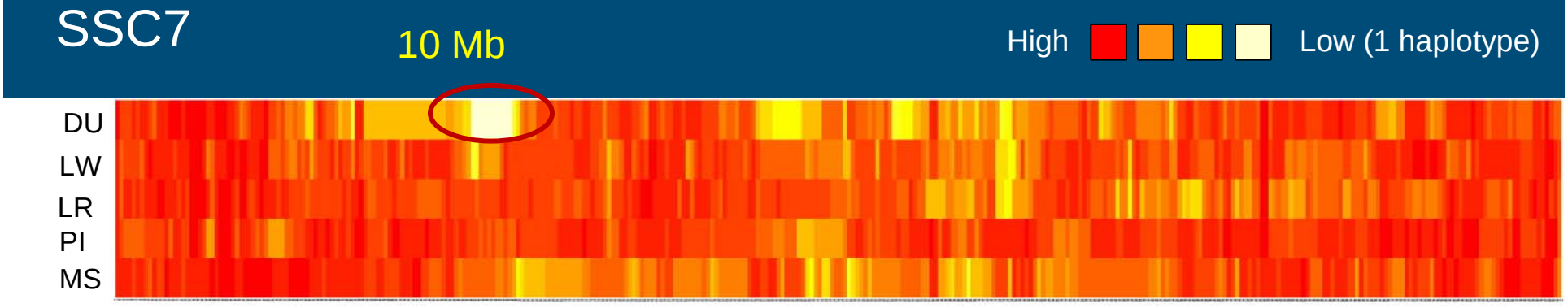
Large W/Landrace
Large W/Pietrain



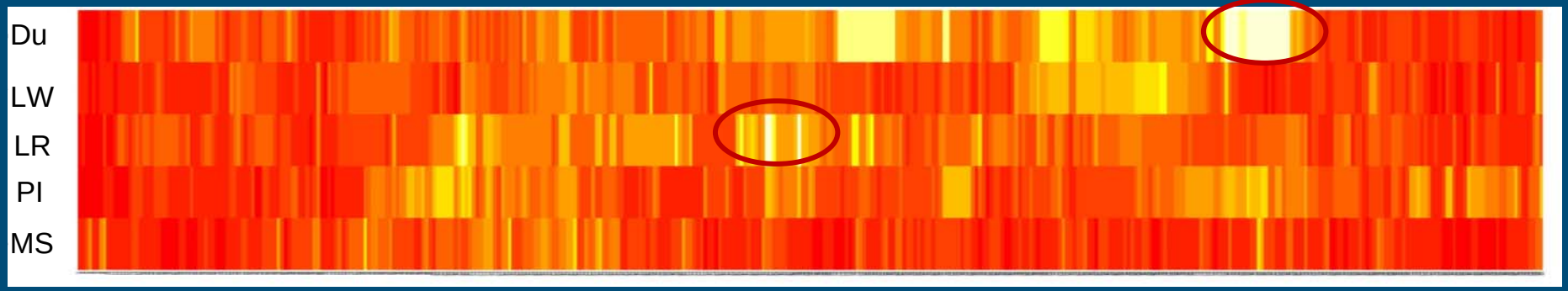
3567 SNPs on SSC7, nearly complete, 133 Mb

Haplotype diversity and linkage disequilibrium

SSC7

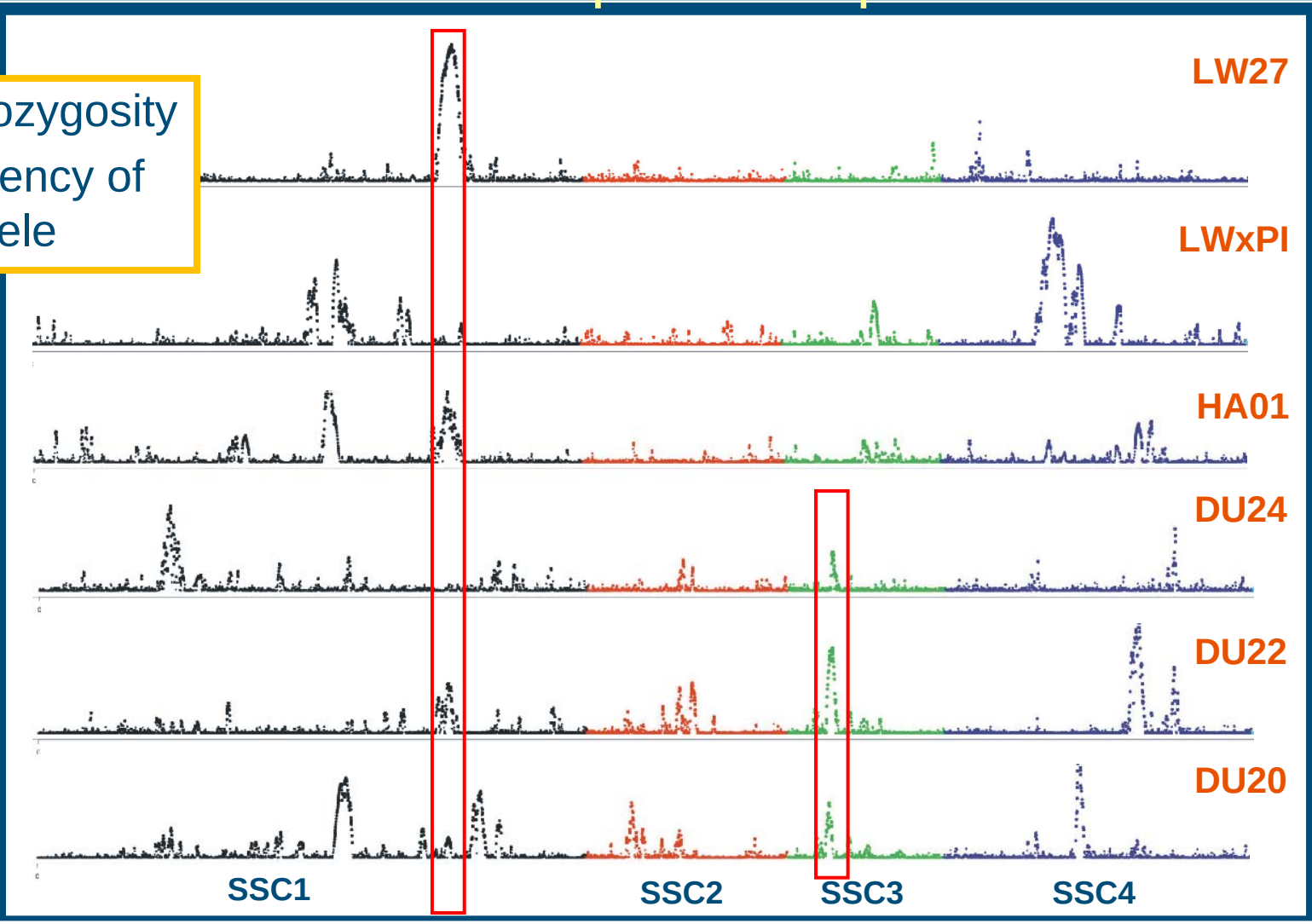


SSC14



Identification of selective sweeps: Examples on SSC1-4

- Low heterozygosity
- High frequency of derived allele



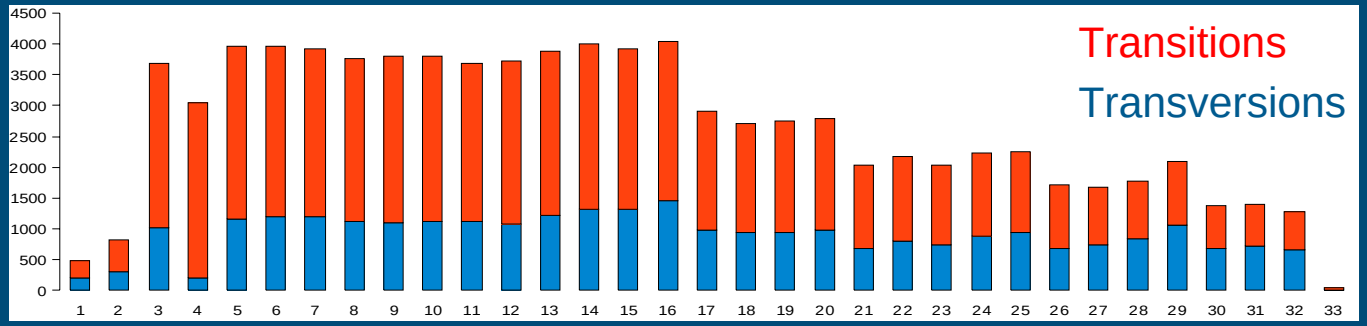
From SNP typing to resequencing

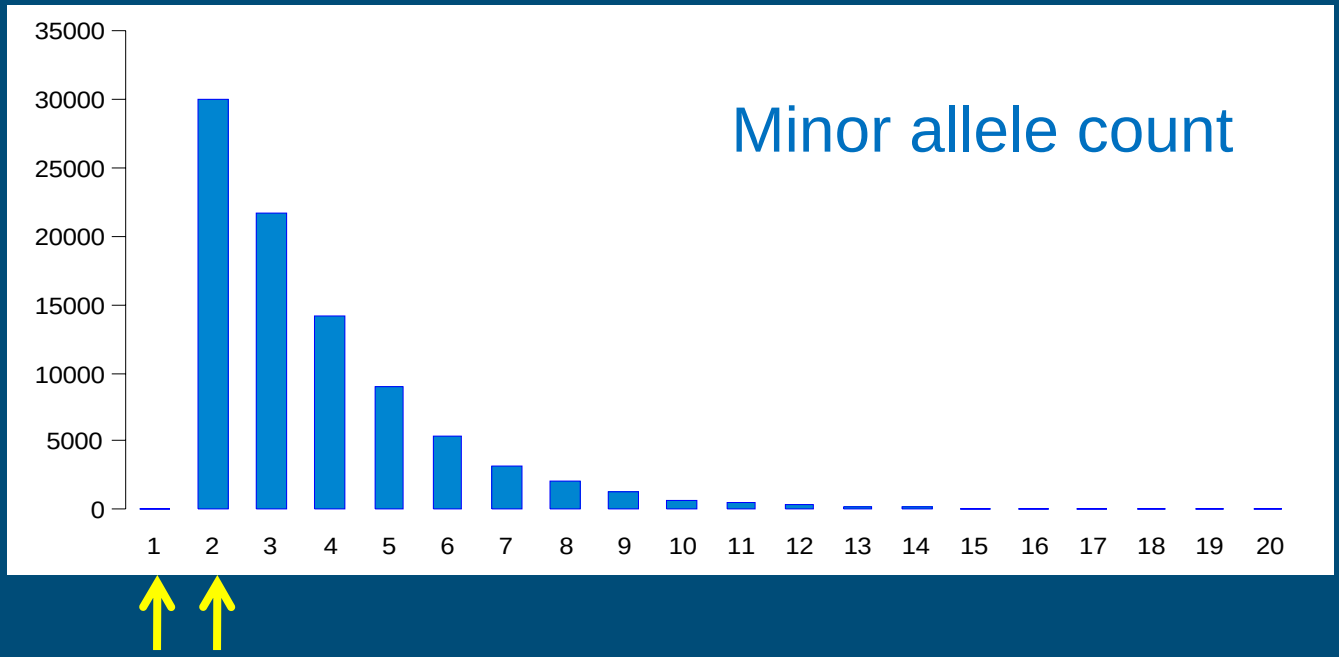
(1) Site frequency spectrum analysis based on RRL sequence data

- RRL cover 5-10 % of the genome
- Coverage differs between breeds
- 4 commercial white breeds + Wild Boar
- Data from pools
- Within 500 Kb windows estimate:
  - Watterson's estimator: θ = f(S, n).
  - Tajima's estimator: π = f(S, n, freqs.).
  - Tajima's D: D = (π-θ) / sd.
  - Fst measures population differentiation

Reanalyse data per breed

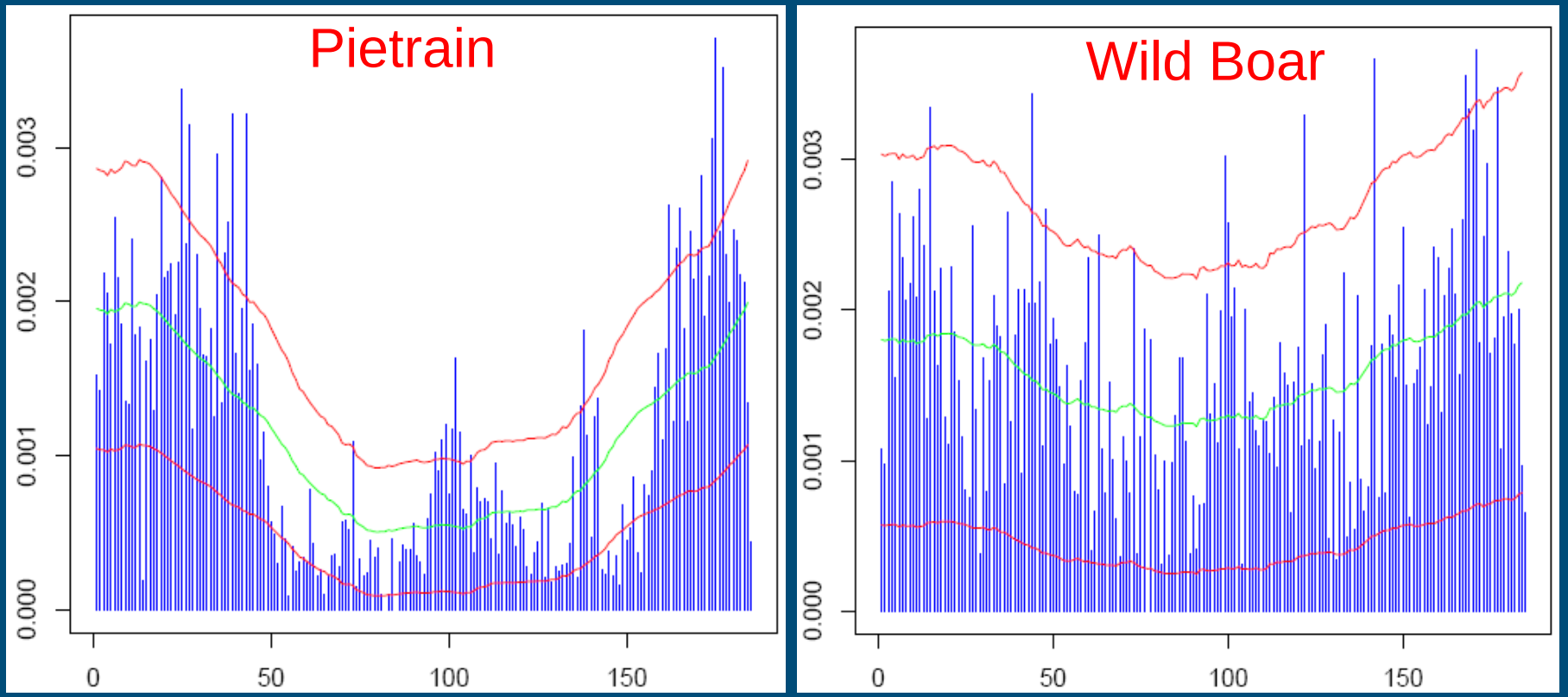
SNP distributions





Example: Site frequency spectrum of Ssc8 derived from deep sequencing of pools

Variation of nucleotide diversity ($\hat{\theta}_W$)



Selective sweep around the *KIT* locus in white breeds

Sequencing of Sus Verrucosus (Java warty pig)

~2 My

- 6x on Illumina GA
- 200, 500 and 3000 bp libraries

Acknowledgements (1)

Funding

- USDA – CSREES
- EU (Sabre, PigSNP)
- Institute for Pig Genetics (IPG)

Food Quality and Safety

DNA samples

- China Agricultural University
  - Ning Li
- University of Sassari, Italy
  - Massimo Scandura
- Staff Institute, Japan
  - Naohiko Okumura
- INRA-Toulouse, France
  - Alain Duvro
- MARC, USA
  - Gerry Rohrer
- Roslin Institute, UK
  - Alan Archibald
- EU PigBiodiv1
- PTP, Italy
  - Elisabetta Guiffra
- REPROGEN, Australia
  - Jaime Gongora
- Universidade de Lisboa, Portugal
  - Deodália Dias
- Universitat Autonoma Barcelona
  - Miguel Perez-Enciso
- University of Aarhus, Denmark
  - Christian Bendixen
- Durham University, UK
  - Greger Larsen
- Iowa State University, USA
  - Max Rothschild
- Wageningen University, The Netherlands
  - Richard Crooijmans
- University of Illinois, USA
  - Larry Schook
- Aristotle University of Thessaloniki, Greece
  - Costas Triantaphyllidis
- Istituto Zootecnico per la sardegna, Italy
  - Sara Casu
- Research Centre for Biology – LIPI, Indonesia
  - Gono Semiadi
- Hendrix Genetics (HG)
- Institute for pg genetics (IPG)
- Pig Improvement Company (PIC)

Acknowledgements (2)

- Wageningen University, The Netherlands
  - Richard Crooijmans
  - Marcos Ramos
  - Andreia Fonseca
  - Hinri Kerstens
  - Hendrik-Jan Megens
- University of Aarhus, Denmark
  - Christian Bendixen
  - Jakob Hedegaard
- USDA,ARS, MARC, USA
  - Gary Rohrer
  - Tim Smith
  - Dan Nonneman
- INRA , Toulouse, France
  - Denis Milan
- Roslin Institute, UK
  - Alan Archibald
  - Andy Law
- Durham University, UK
  - Greger Larsen
- Purdue University, USA
  - Bill Muir
- University of Illinois, USA
  - Larry Schook
  - Jon Beever
- USDA, ARS, Beltsville, USA
  - Curt Van Tassel
- Iowa State University, USA
  - Max Rothschild
  - Zhiliang Hu
  - James Reecy
- Sanger Institute, UK
  - Carol Churcher
  - Richard Clark
- University of Missour-Columbia, USA
  - Jeremy Taylor
  - Bob Schnabel
- Universitat Autonoma Barcelona
  - Miguel Perez-Enciso
  - Lucca Ferreti
- Illumina Inc., San Diego, USA
  - Mark Hansen
  - Marylinn Munson