



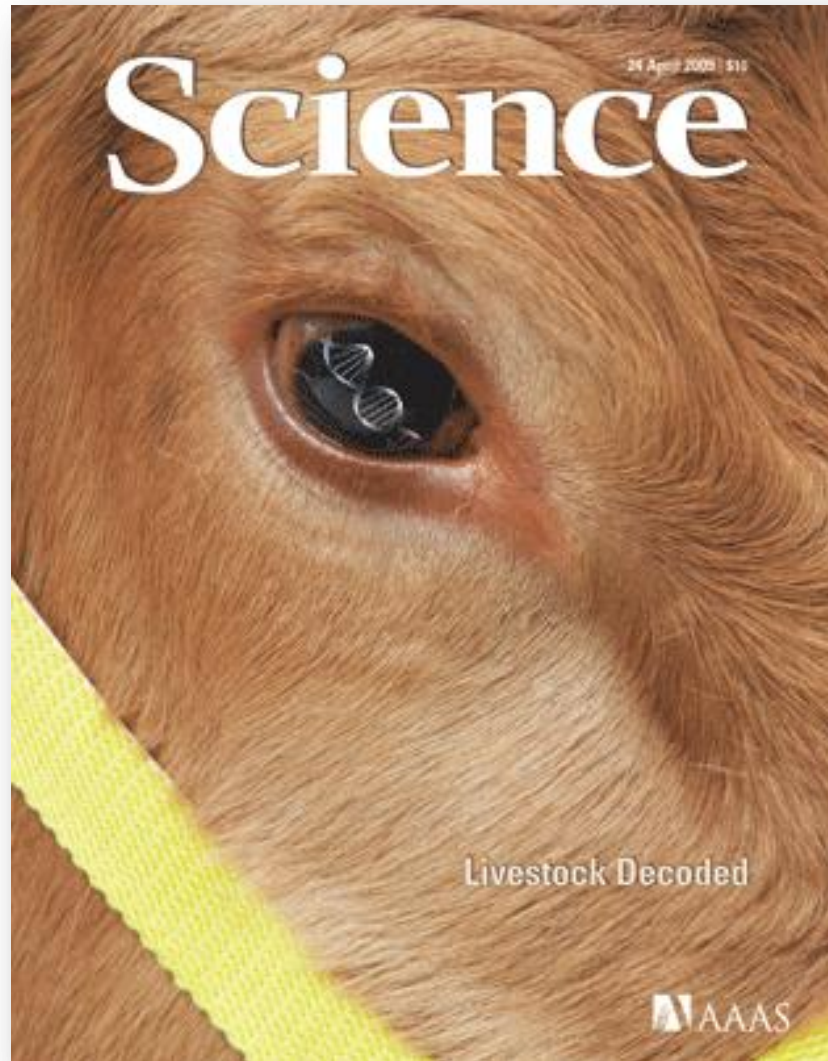
Декодирование геномов для понимания процессов селекции и эволюции у млекопитающих

Denis M. Larkin
Aberystwyth University
Institute of Biological, Environmental and Rural Sciences

Sequencing of Genomes

- ▶ Adaptation of individual species/genomes to environment (cattle, yak, pig)
 - ▶ Natural selection and demography in pigs
 - ▶ Artificial selection (dairy cattle)
-

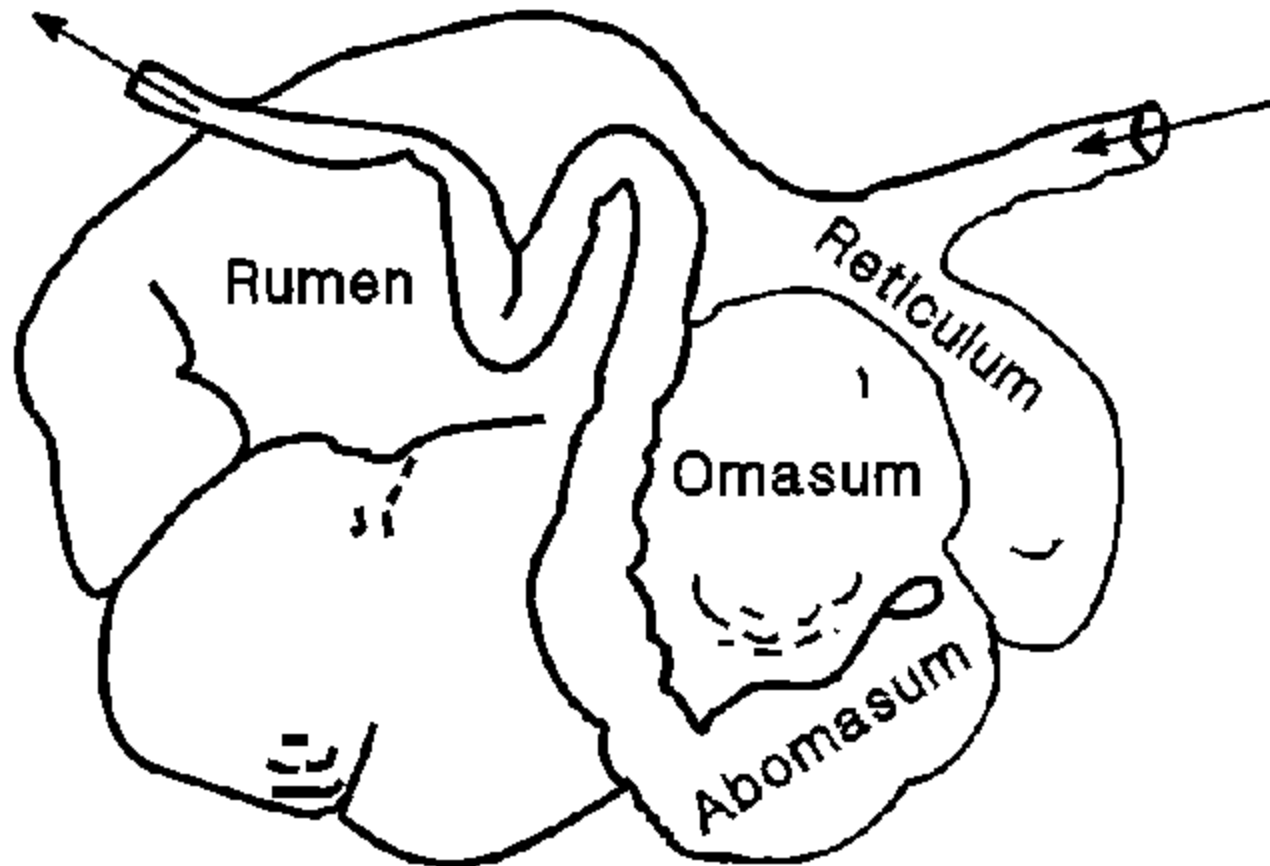
Analysis of the Cattle Genome



Cattle Genome Statistics

Statistics	Value
Coverage	6-7 X
N50 (scaffolds)	1.9 Mbp
Total length	2,870 Mbp
No. genes	22,000
Chromosome assembly	Yes

Rumen

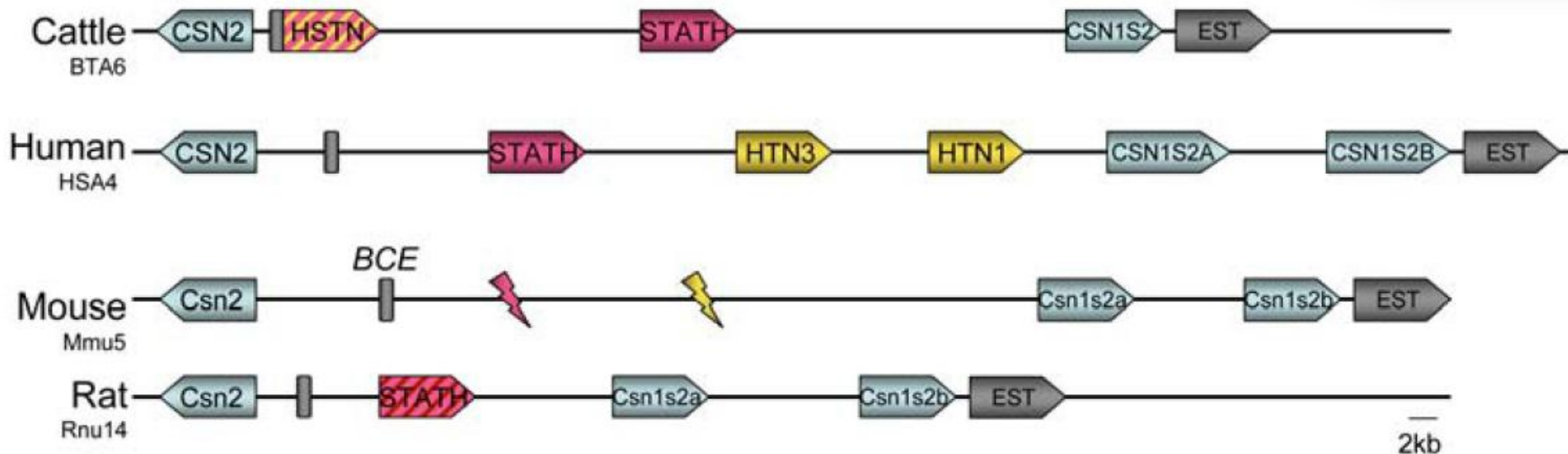




Lineage-Specific Biology

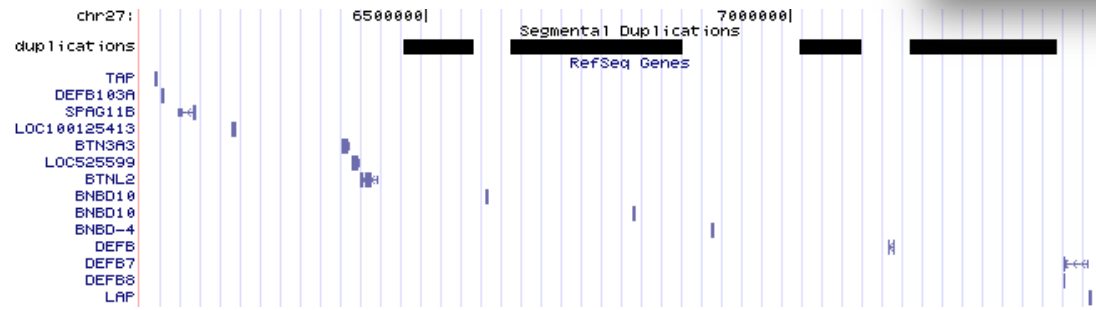
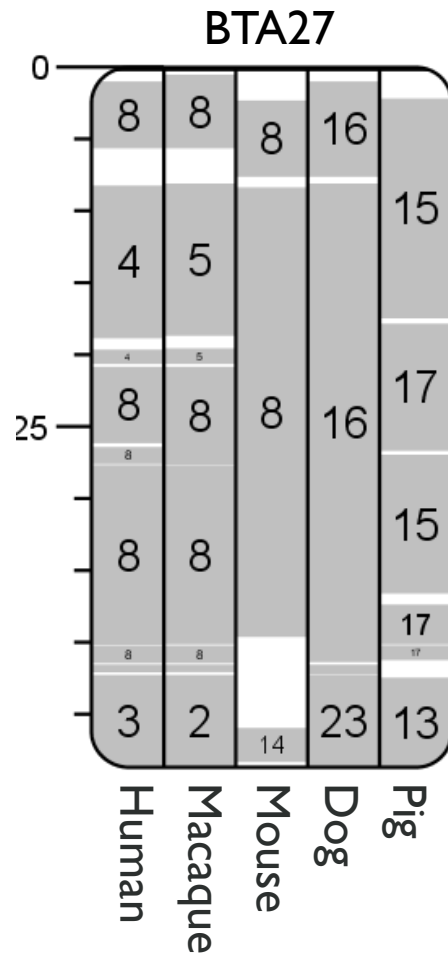
- ▶ Presence of a large number of microbes in the rumen
- ▶ Expansion of gene families related to immune response
 - ▶ β -defensin genes
 - ▶ cathelicidin genes
 - ▶ C-class lysozyme genes
- ▶ Reorganization of gene families encoding proteins present in milk
 - ▶ histatherin (HSTN)

HSTN Gene is Juxtaposed to a Regulatory Element (BCE) in Cattle





Beta-Defensin Cluster in Cattle Is Located in a Cetartiodactyl Evolutionary Breakpoint Region



Gene Family	Cattle	Human	Mouse
Beta-defensin	106	39	52

Yak Genome & Evolution

The yak genome and adaptation to life at high altitude

Qiang Qiu^{1,16}, Guojie Zhang^{2,16}, Tao Ma^{1,16}, Wubin Qian^{2,16}, Junyi Wang^{2,16}, Zhiqiang Ye^{3,4,16}, Changchang Cao², Quanjun Hu¹, Jaebum Kim^{5,6}, Denis M Larkin⁷, Loretta Auvil⁸, Boris Capitanu⁸, Jian Ma^{5,9}, Harris A Lewin¹⁰, Xiaojun Qian², Yongshan Lang², Ran Zhou¹, Lizhong Wang¹, Kun Wang¹, Jinqian Xia², Shengguang Liao², Shengkai Pan², Xu Lu¹, Haolong Hou², Yan Wang², Xuetao Zang², Ye Yin², Hui Ma¹, Jian Zhang¹, Zhaofeng Wang¹, Yingmei Zhang¹, Dawei Zhang¹, Takahiro Yonezawa¹¹, Masami Hasegawa¹¹, Yang Zhong¹¹, Wenbin Liu², Yan Zhang², Zhiyong Huang², Shengxiang Zhang¹, Ruijun Long¹, Huanming Yang², Jian Wang², Johannes A Lenstra¹², David N Cooper¹³, Yi Wu¹, Jun Wang^{2,14}, Peng Shi³, Jian Wang² & Jianquan Liu^{1,15}

Adapted to live at high altitudes (hypoxia)

Genome sequenced using Illumina technology

Genome assembled to ~2,000 large scaffolds

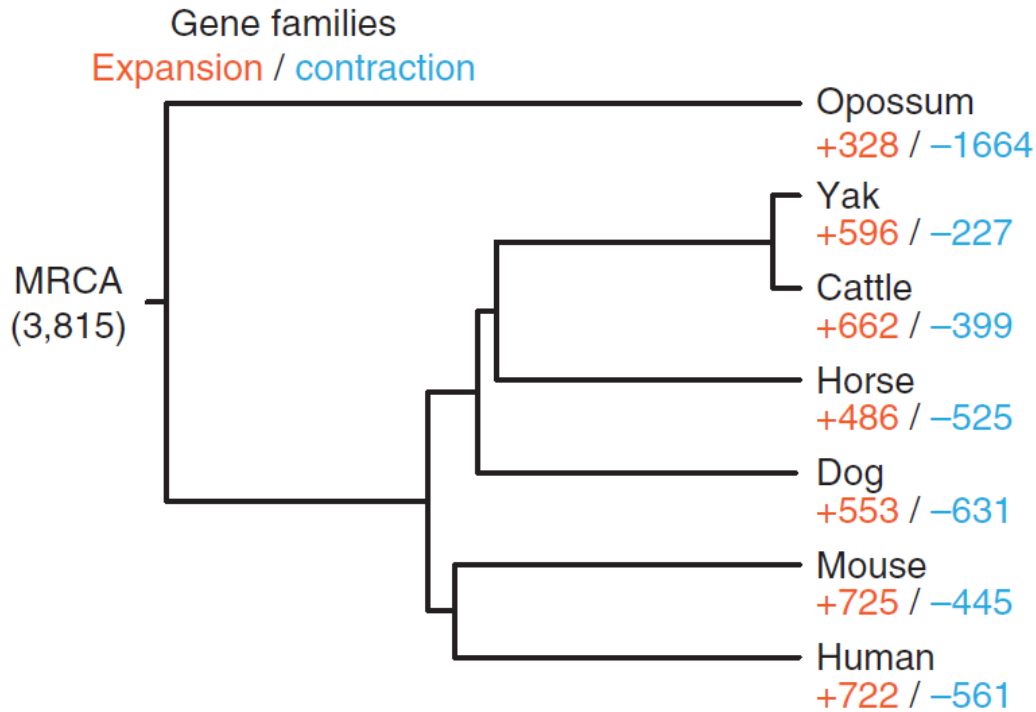
No chromosome assembly available



Yak Genome Statistics

Statistics	Value
Coverage	65 X
N50 (scaffolds)	1.4 Mbp
Total length	2,657 Mbp
No. genes	22,282
Chromosome assembly	No

Yak Gene Family Expansions



Functional categories (P-value <0.01)

Sensory perception

Sensory perception of taste

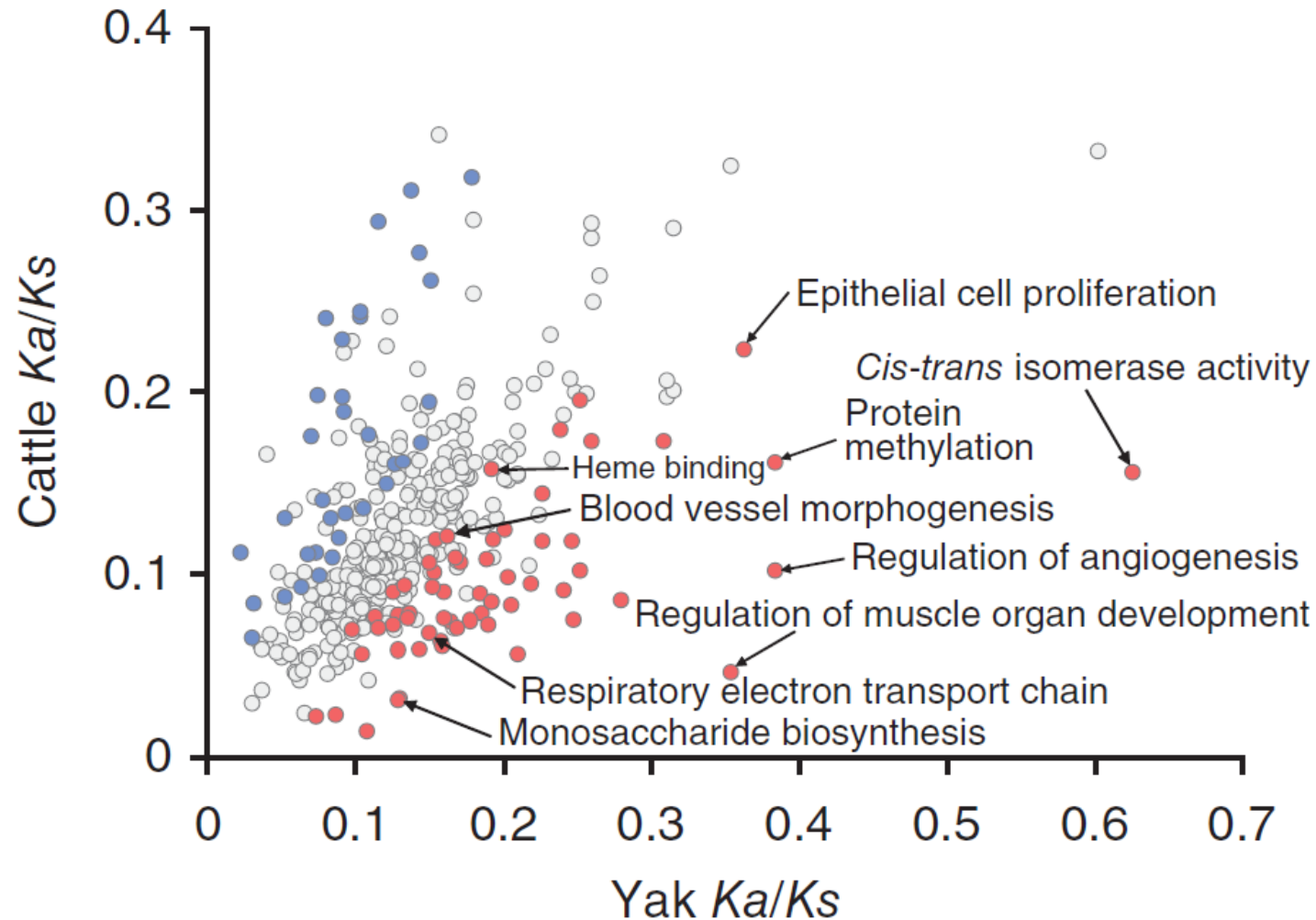
Olfactory receptor activity

Energy metabolism

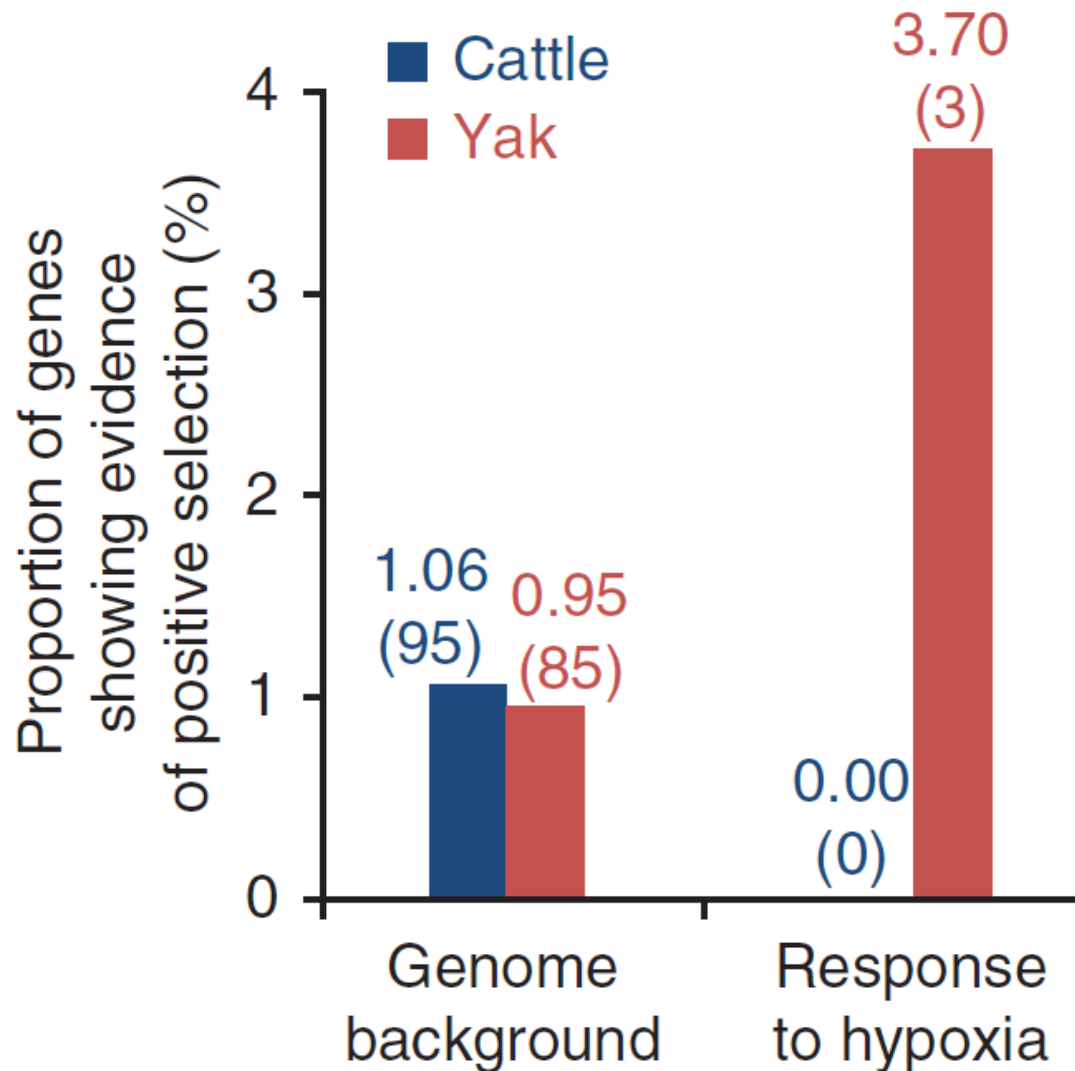
Cytochrome-c oxidase activity

ATP synthesis coupled proton transport

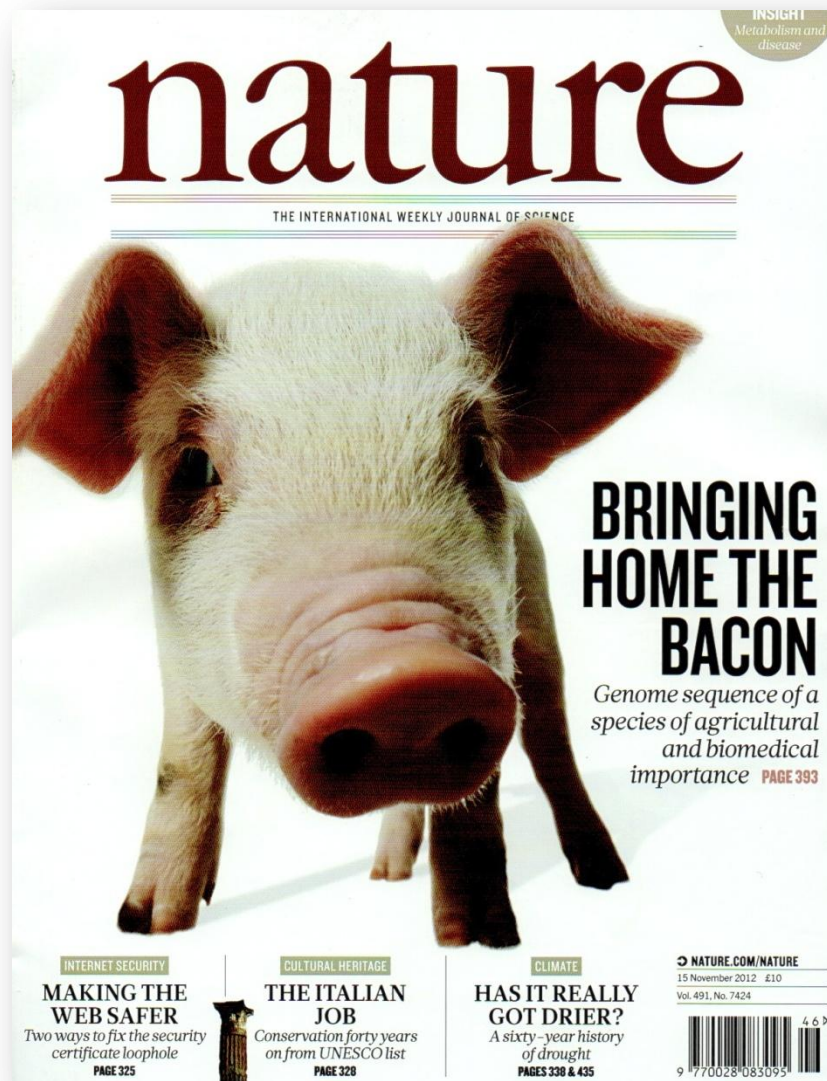
Adaptive Evolution in the Yak Lineage



Positive Selection in the Yak and Cattle Lineages



Analysis of the pig genomes



Pig Genome Statistics



Statistics	Value
Coverage	~30 X
N50 (scaffolds)	0.64 Mbp
Total length	2,597 Mbp
No. genes	21,640
Chromosome assembly	Yes

Immune Gene Family Expansions

Pig-specific

Type I Interferon, δ -subfamily

Cetartiodactyl-specific

Cathelicidin superfamily

Type I Interferon (inclusive)

Ruminant-specific

β -defensin superfamily

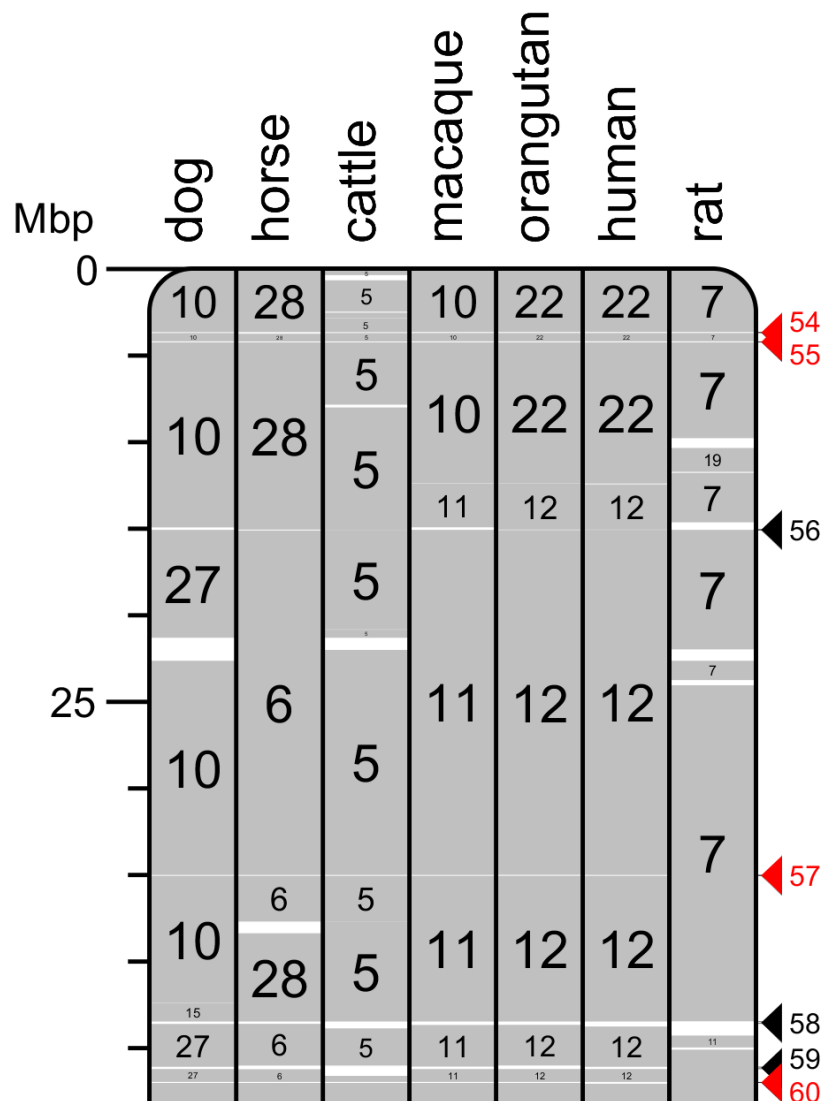
C-type Lysozyme/LYZ1 superfamily

Type I Interferon, β -subfamily

Type I Interferon, τ -subfamily



Detection of Gene Networks Affected by Genomic Rearrangements in Pigs



SSC10

Resolution (Kb)	Pig EBRs	Artiodactyl EBRs
100	1,006	77
300	193	20
500	146	16
Consensus	192	20

IPGSC, *Nature*, November 15, 2012



Gene Functional Process Networks in Pig EBRs

Processes	P-values	Ratio
Sensory perception of taste	8.9e-6*	11/23
Glutathione metabolic process	8.0e-4	9/25
Sensory perception of bitter taste	1.3e-3	5/9
Midbrain-hindbrain boundary development	1.3e-3	5/9
Regulation of protein ubiquitination involved in ubiquitin-dependent protein catabolic process	1.3e-3	5/9

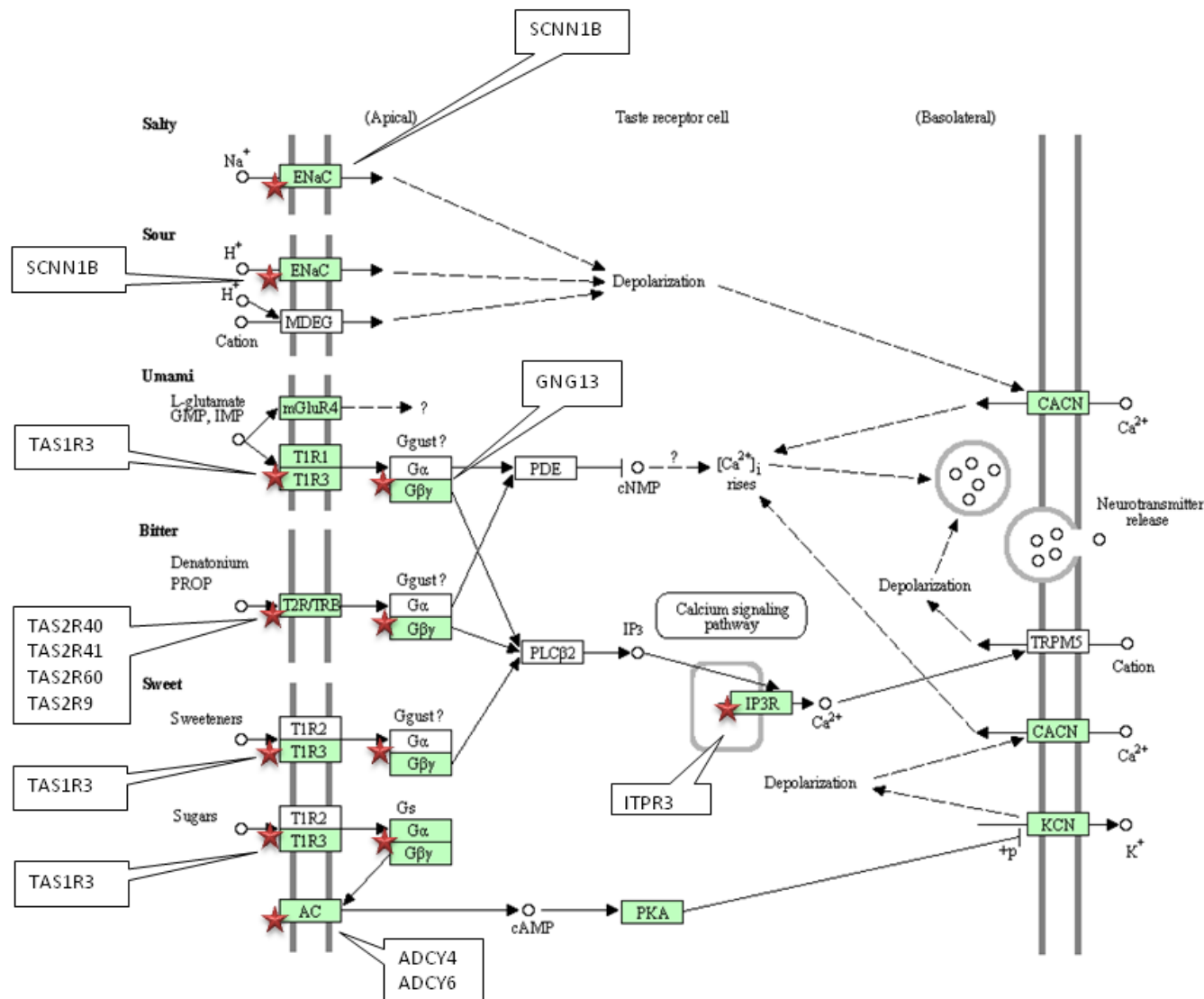
*FDR < 0.05

Taste Perception Genes in Pig EBRs

EBR coordinates (Mbp)	Genes
SSC1: 0.30-0.30	DBH
SSC3: 41.57-41.62	GNGI3
SSC5: 15.06-15.06	ADCY6,WNT10B
SSC5: 63.74-63.79	TAS2R9
SSC6: 57.78-57.81	TAS1R3
SSC7: 34.13-34.13	ITPR3
SSC7: 79.94-79.94	ADCY4
SSC10: 0.34-0.39	SCNN1B
SSC18: 6.77-6.82	TAS2R41,TAS2R60,TAS2R40
SSC18: 53.34-53.40	NPY



Taste Perception Network is Affected by Chromosome Rearrangements in Pigs



Perception of Salty Tastes

- ▶ Pigs have low ability to taste salty compounds (Danilova et al., 1998)
- ▶ Key genes: SCNN1A, SCNN1B, SCNN1G
- ▶ Reposition and probable duplication of SCNN1B gene in pigs



Perception of Bitter Tastes in Pigs

- ▶ Pigs react to higher concentrations of bitter compounds than humans
 - ▶ *Denatonium benzoate*: pigs react to concentrations 10,000x higher than those rejected by humans
- ▶ Human genome contains 24 bitter taste receptors
 - ▶ Two clusters: HSA7 and HSA12
- ▶ Pig genome contains 17 bitter taste receptors
 - ▶ Two clusters: SSC18 and SSC5
 - ▶ Both clusters are affected by pig-specific rearrangements



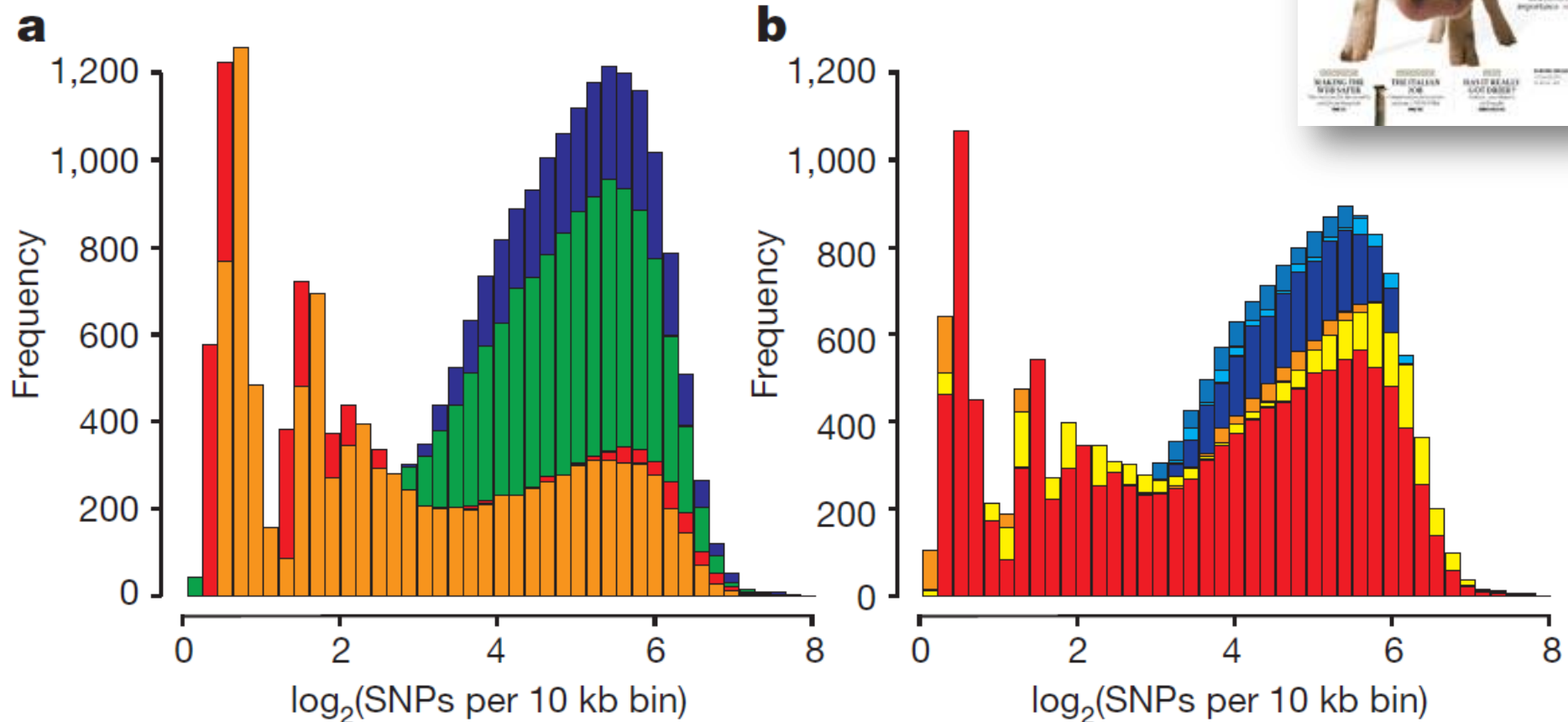
Pigs can consume food not suitable to humans
A reason for domestication?

Population Divergence and Domestication

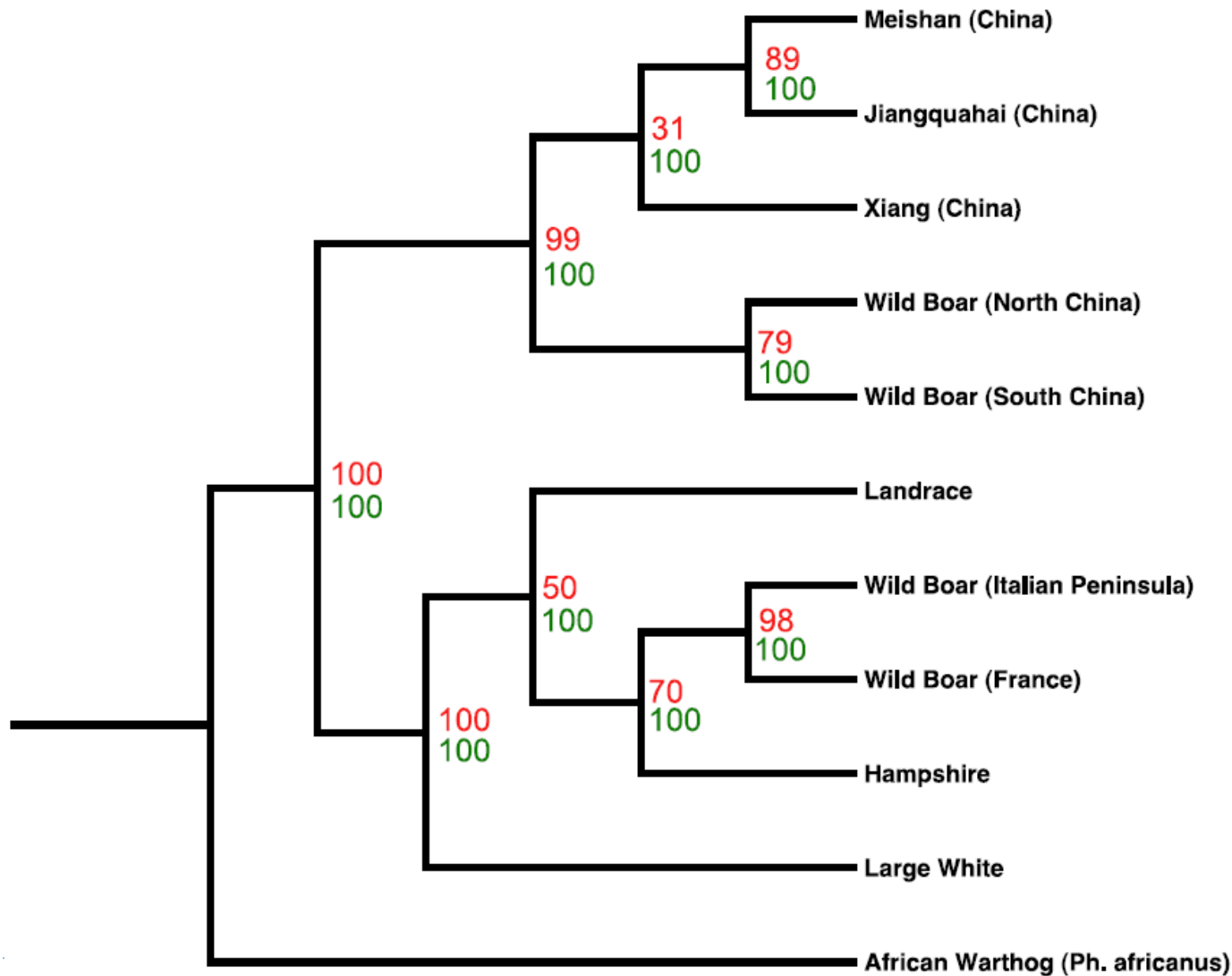
- ▶ 10 wild boars from different areas were sequenced
 - ▶ 17,210,760 single nucleotide polymorphisms detected
 - ▶ 11,472,192 were segregating in 4 Asian boars
 - ▶ 6,407,224 were segregating in 6 European boars
 - ▶ 2,212,288 were shared between Asian & European boars



Distribution of Heterozygosity for Individual Pig Genomes



Phylogenetic Relation Between Sequenced Pigs



Artificial Selection (Dairy Cattle)

Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle

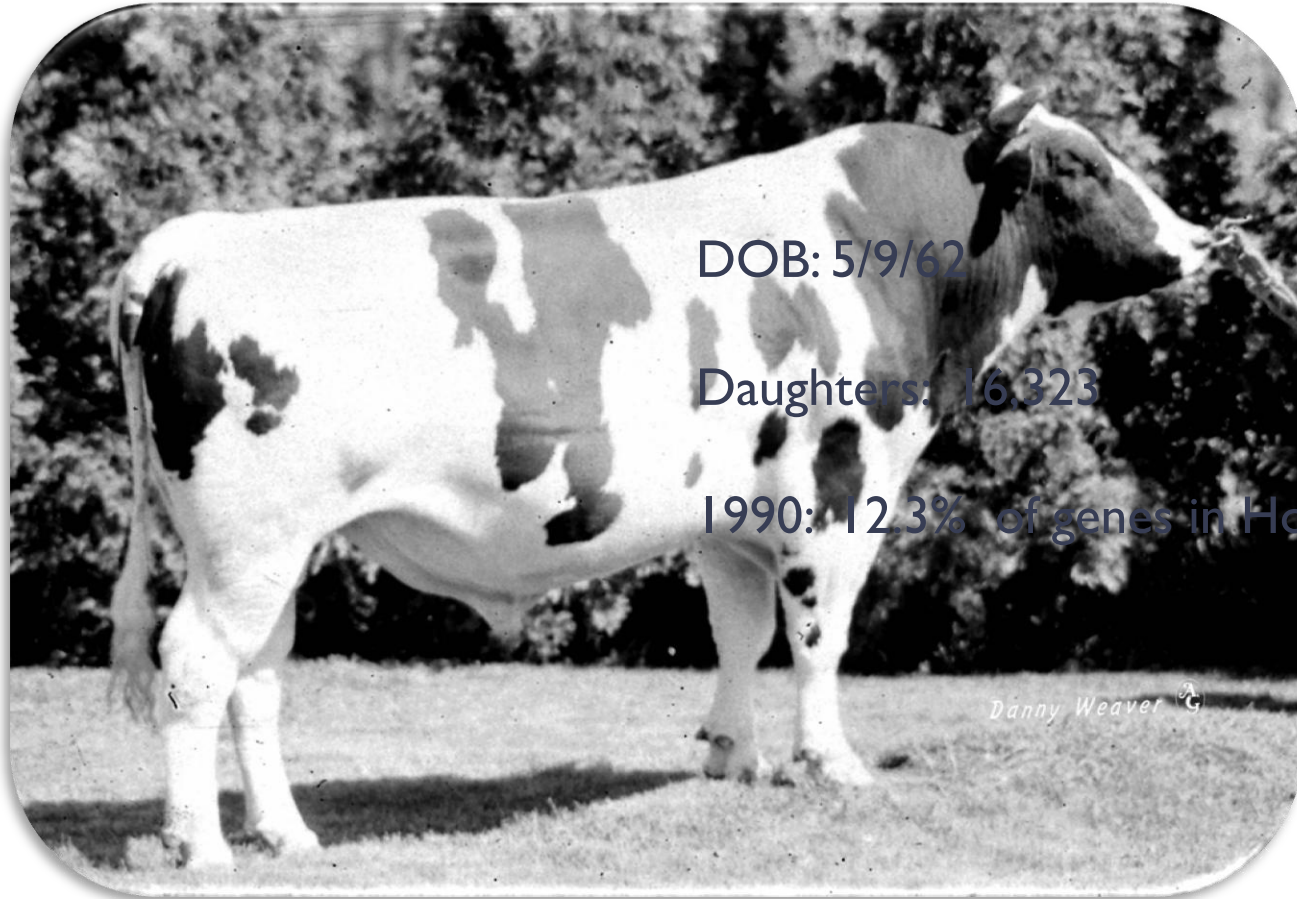
Denis M. Larkin^a, Hans D. Daetwyler^b, Alvaro G. Hernandez^c, Chris L. Wright^c, Lorie A. Hetrick^c, Lisa Boucek^c, Sharon L. Bachman^c, Mark R. Band^c, Tatsiana V. Akraiko^c, Miri Cohen-Zinder^d, Jyothi Thimmapuram^c, Iona M. Macleod^e, Timothy T. Harkins^f, Jennifer E. McCague^g, Michael E. Goddard^{b,e}, Ben J. Hayes^{b,h}, and Harris A. Lewin^{a,d,1,2}

^aDepartment of Animal Sciences, ^bThe W. M. Keck Center for Comparative and Functional Genomics, and ^dInstitute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ^cBiosciences Research Division, Department of Primary Industries, Bundoora 3083, Victoria, Australia; ^eDepartment of Food and Agricultural Systems, University of Melbourne, Parkville 3011, Victoria, Australia; ^fRodhe, Indianapolis, IN 46250; ^g454 Life Sciences, Branford, CT 06405; and ^hBiosciences Research Centre, La Trobe University, Bundoora 3086, Victoria, Australia

Walkway Chief Mark



Pawnee Farm Arlinda Chief



DOB: 5/9/62

Daughters: 16,323

1990: 12.3% of genes in Holsteins

Danny Weaver ©

Why Sequence Chief and Mark?

- ▶ Chief and Mark contributed a significant fraction of the genome in the current Holstein population
- ▶ Mark's sons were previously evaluated for QTL using the granddaughter design (Heyen et al. 1999; Ashwell et al. 2004; 2005)
- ▶ All of the current top 10 sires for milk yield have Chief in their pedigree (AIPL)

Project Goal

- ▶ Simultaneous identification of genetic polymorphisms responsible for multiple production and conformation traits in Holsteins using whole-genome sequencing combined with genome-wide association data

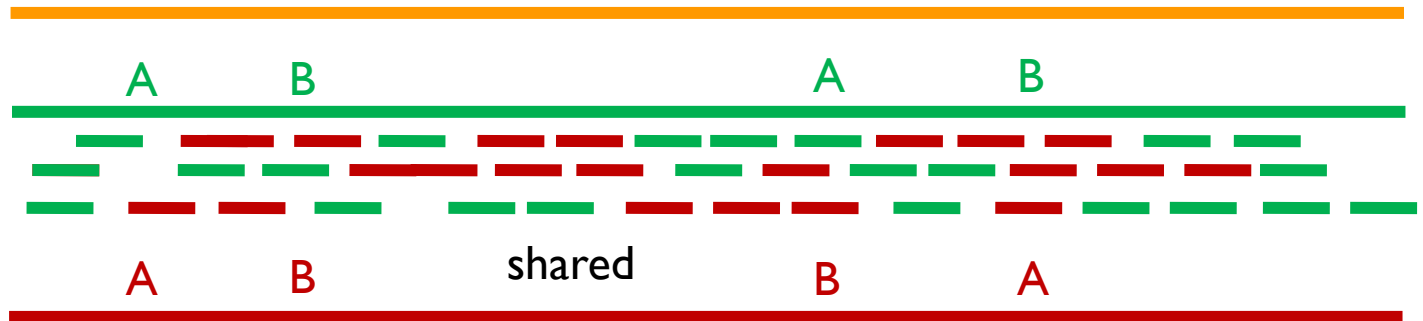
Approach

- ▶ Whole genome resequencing (Roche 454 FLX Titanium) of Chief (6x) and Mark (12x)
- ▶ Genome-wide reconstruction of Mark haplotypes inherited from Chief
- ▶ Track Chief's haplotypes through successive generations using SNP genotyping data
- ▶ Identify QTL-bearing Chief haplotypes in current population using GWAS data
- ▶ Fine map QTL and indentify candidate Chief QTNs

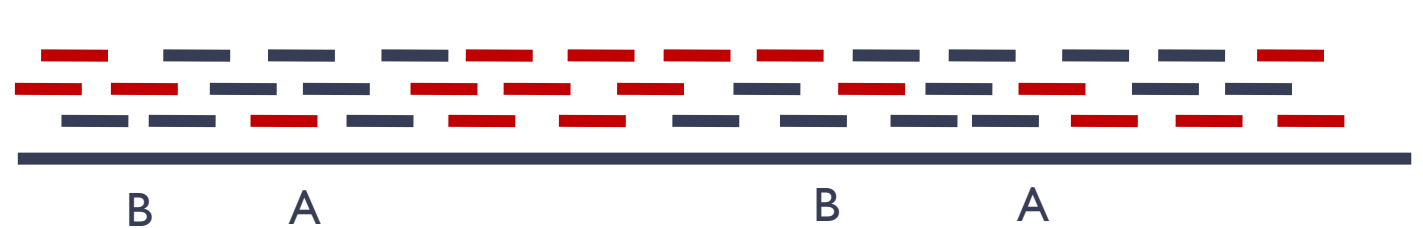
Haplotype Reconstruction by Resequencing

Btau4.0

6x

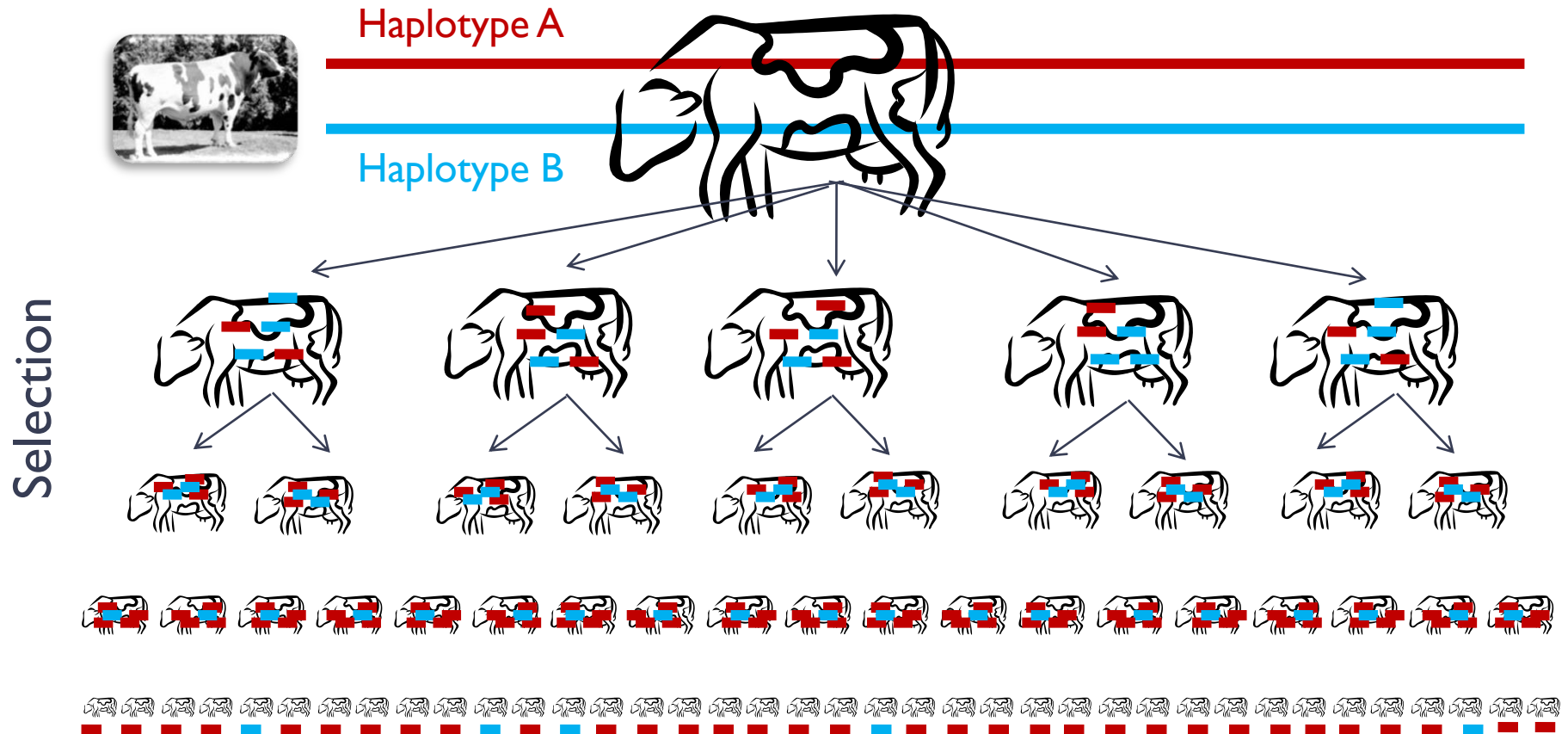


12x



Probability of correctly identifying shared allele at SNP locus is $> .988$

Haplotype Tracking & Fine Mapping QTL (sequencing + dense SNP genotyping + association)



SNP Detection Statistics & Filtering Criteria

SNP Set	“Homozygous”	“Heterozygous”
Initial SNP set	1,851,126	10,583,734
Homopolymer regions	1,672,815 (90.4%)	6,992,325 (66.1%)
SNPs with reads from a single direction	1,247,978 (67.4%)	2,787,407 (26.3%)
Remaining set	1,207,103 (65.2%)*	1,356,094 (12.8%)**
Haplotype phase reconstruction	N/A	972,479 (9.2%)

* SNPs vs. reference defined by ≥ 3 reads for new allele

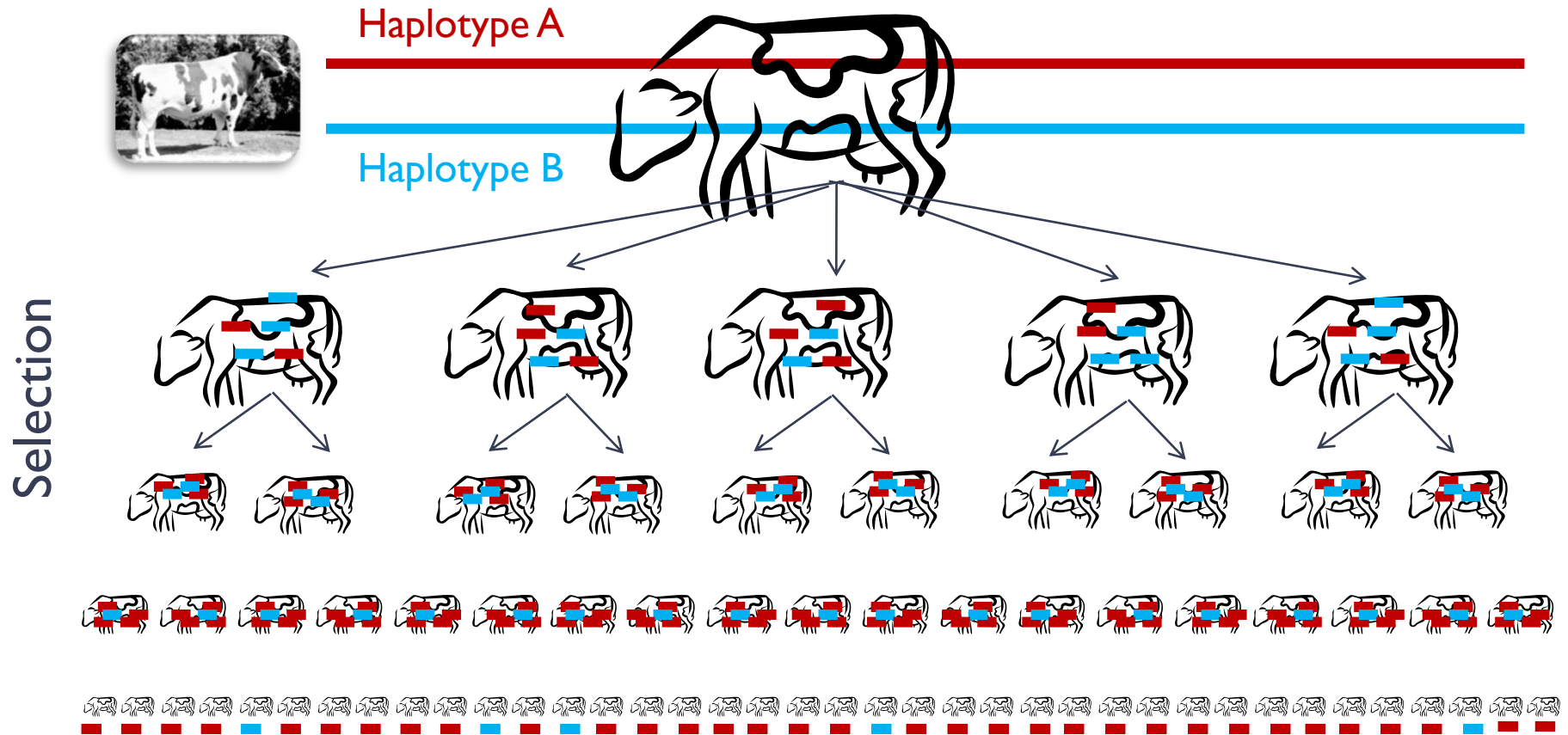
** SNPs heterozygous in sires defined by ≥ 3 reads for new allele in Mark, ≥ 2 reads in Chief

Agreement in Allele Definitions for Haplotypes Defined by 454 Sequencing and SNP50 Genotyping

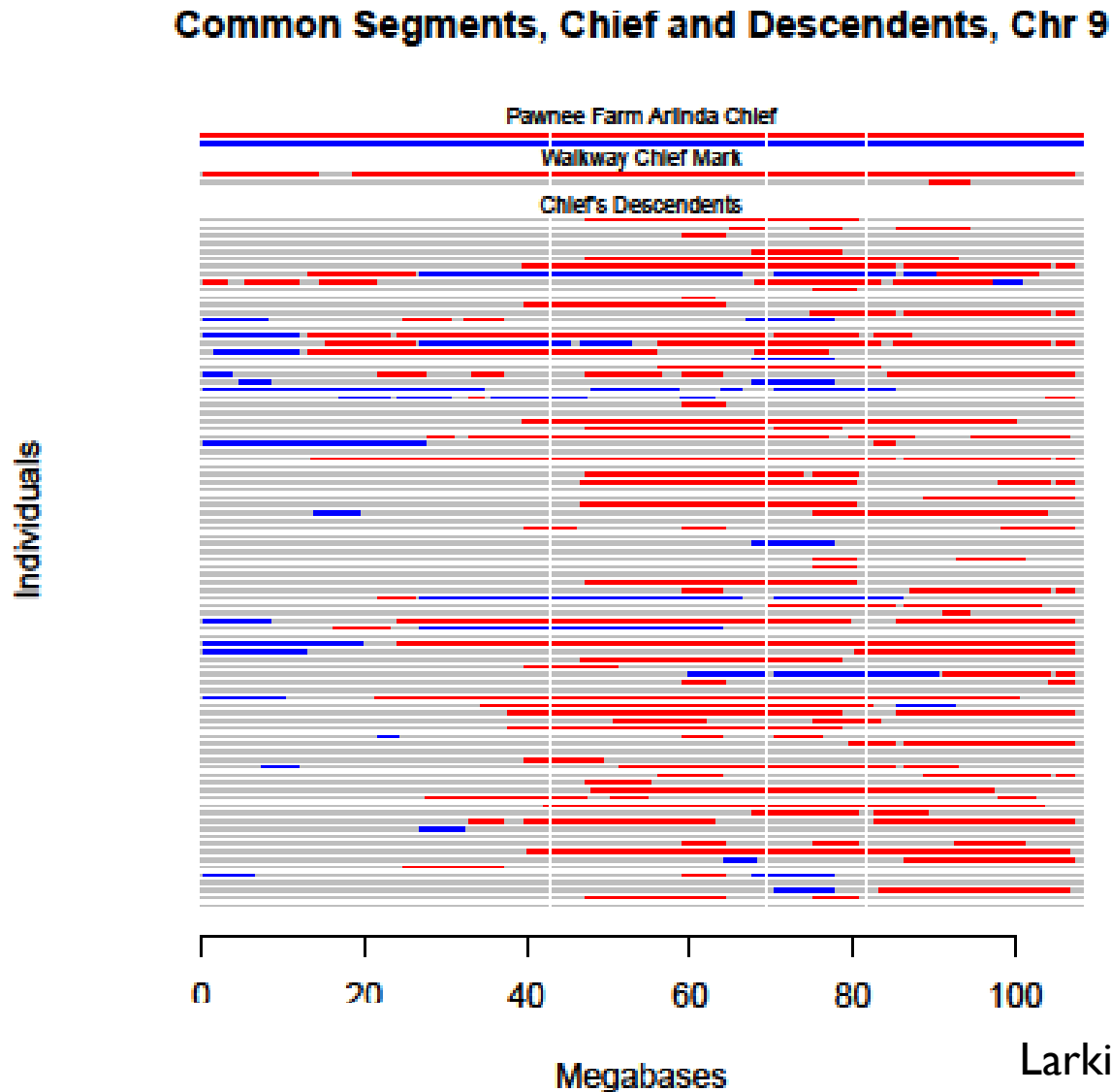
Min. read coverage	No. SNPs	No. SNP50 SNPs*	Agreement (%)
3	1,408,303	8,196	89.0
6	505,043	3,336	95.8
7	368,754	2,495	96.8

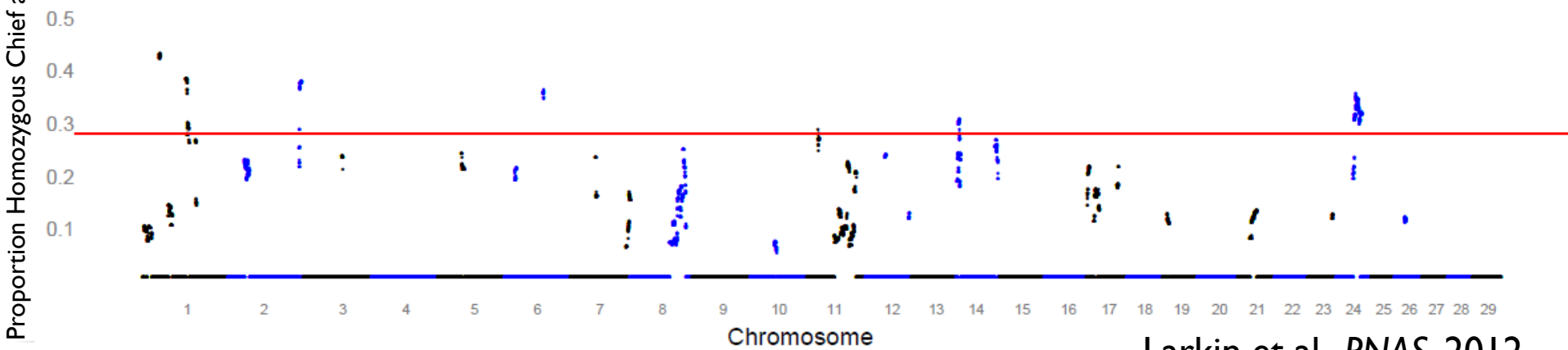
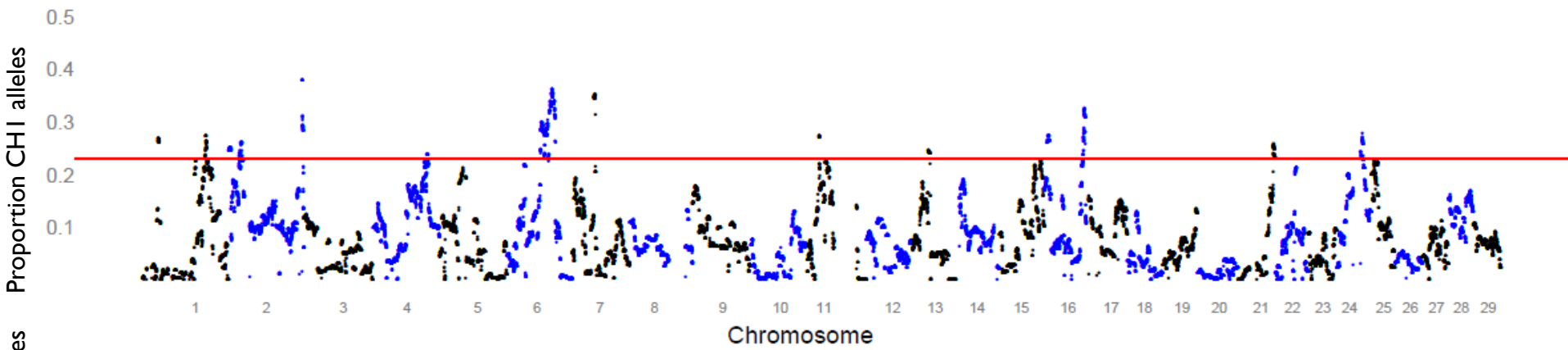
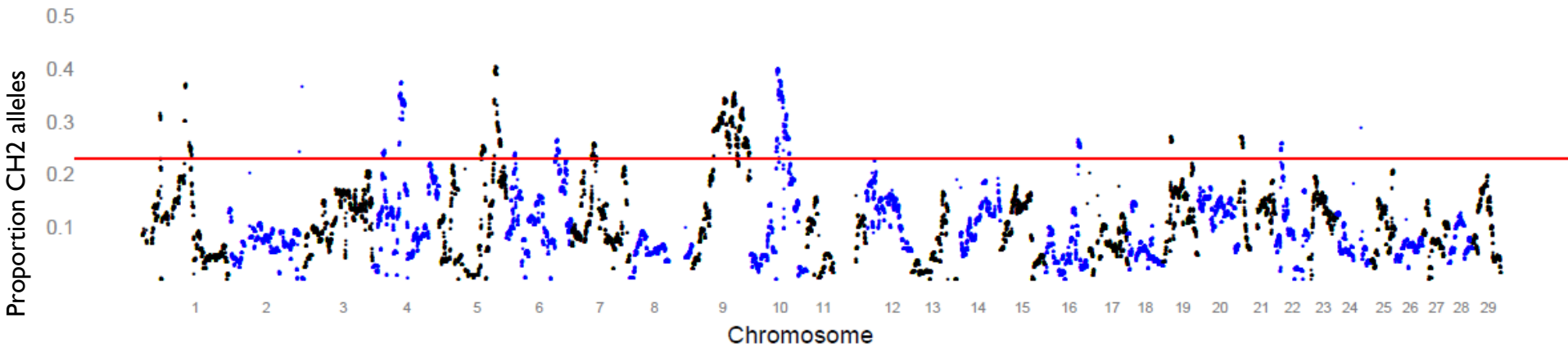
* Based on Illumina SNP50 genotyping of Mark and 92 of his DBDR offspring

Haplotype Tracking & Fine Mapping QTL (sequencing + dense SNP genotyping + association)

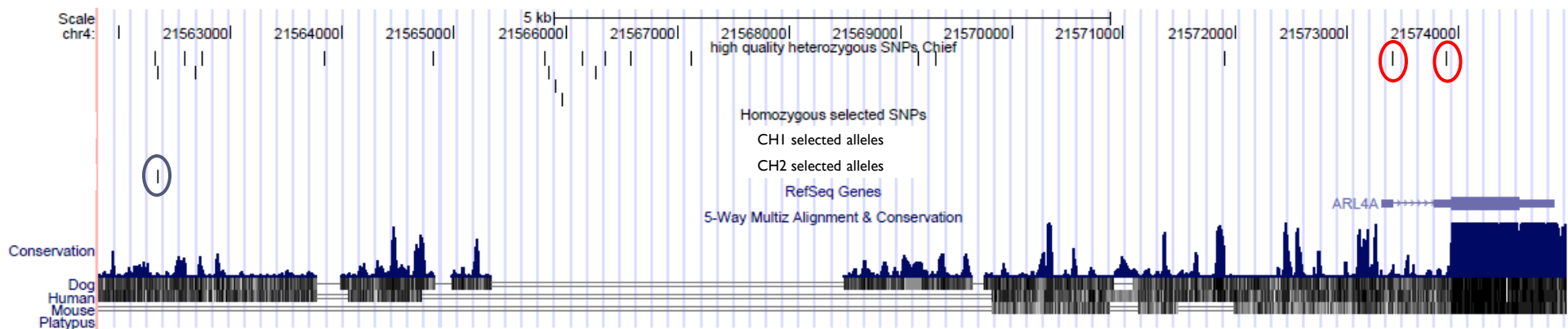


Segments in common between Chief and 50 of his direct descendents

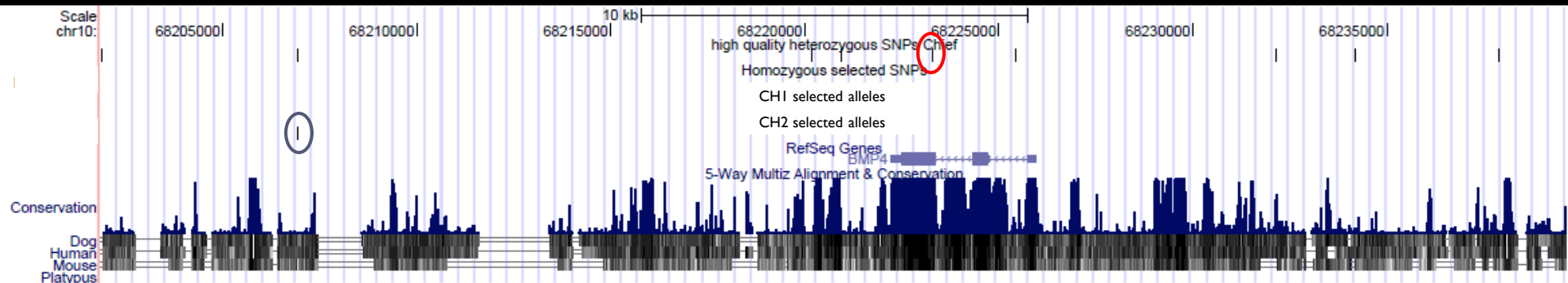




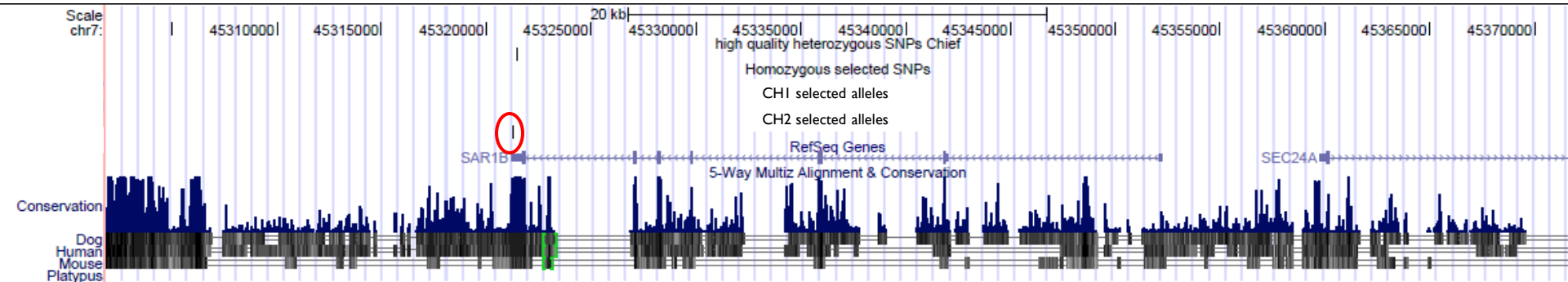
ARL4A



BMP4



SAR1B



Summary & Conclusions

- ▶ Genome sequencing is a power tool to understand the underlying genetics of lineage- or species-specific biology
 - ▶ Scenery perception gene networks are the commonly affected by the process of adaptation/or speciation
 - ▶ Resequencing multiple genomes from the same species allow for an effective detection of selective sweeps and demographic history
 - ▶ Resequencing 'right' genomes in the populations subjected to intensive (artificial) selection allows for detection of the selected regions and the underlying mutations
-

Collaborators & Acknowledgments



Jaebum Kim
Jian Ma



Harris Lewin



Guoji Zhang
Qingle Cai
BGI Sequencing Team



Loretta Auvil
Boris Capitanu
Mike Welge



Dynasty Foundation
