Whole genome sequencing of six dog breeds from continuous altitudes reveals adaption to high-altitude hypoxia

Yixue Li Ph.D

yxli@sibs.ac.cn

Key Lab of Systems Biology
Shanghai Institutes for Biological Sciences
Chinese Academy of Sciences

"For such a large number of problems there will be some animal of choice, or a few such animals, on which it can be most conveniently studied"

August Krogh et al., Am J Physiol. 90(2) pp. 243-251(1929)

August Krogh: *a* Danish Nobel Laureate and a physiologist, he got his Nobel praise last century on 1920.

Background

Chronic mountain sickness

- In general, if lowlanders just moved to highland, at the first few days, they got altitude stress because of the low capacity of oxygen uptake, after then the hemoglobin concentration in blood will be increased during acclimatization compensates for the unavoidable lowered percent of oxygen saturation (O2 sat) of Hb due to ambient hypoxia.

- Lowlanders residing at altitudes higher than 2500 meters (m) are at risk for chronic health problems arising in part from acclimatization processes. For example, long-term high Hb levels increase blood viscosity as well as the risk of thrombosis and stroke and poorer pregnancy outcomes. These results taken together suggest that the acclimatization response does not assure that fitness is unaltered at High altitude.

Scientific interests and questions

- ***A short-term adaptive response to living at high altitudes, can cause excessive production of red blood cells, leading to chronic mountain sickness that can kill people or reduce their reproduction.***
- ***But the facts one found that people living permanently at high altitudes might undergo selection at some genes to reduce the fitness consequences of excessive red blood-cell production or erythrocytosis.***
- ***The mechanisms of organism adaptation to high-altitude hypoxia are very interesting questions and are still unclear.***

- The increased oxygen uptake and delivery are physiological hallmarks of high-altitude adaption, and we know that the capacity of oxygen uptake is determined by hemoglobin concentration and oxygen affinity.

- Recent study has also shown that reprogramming cells to pluripotency requires a shift from oxidative to glycolytic metabolism, and requires hypoxia-inducible factors (HIFs) in a stage specific manner. (J. Mathieu et.al., ***Cell***, 2014)

As erythrocytosis is a common symptom of chronic mountain sickness which will lead to high blood viscosity and cardiovascular disorders. Whereas, the decrease in hemoglobin level may provide a protective mechanism for people live in highland.

- Whole-genome genotyping and re-sequencing have been performed for three typical highland populations including ***Tibetans*** (Beall et al. 2010; Bigham et al. 2010; Simonson et al. 2010; Yi et al. 2010; Peng et al. 2011; Xu et al. 2011), ***Andeans*** (Bigham et al. 2009; Bigham et al. 2010) and ***Ethiopians*** (Alkorta-Aranburu et al. 2012; Scheinfeldt et al. 2012).

Tibetans maintain a nearly normal level of hemoglobin concentration and a low level of oxygen saturation, but they display a high level of blood flow, resulting in the increase of oxygen delivery (Beall et al. 2001; Erzurum et al. 2007).

Physiological hallmarks among people

- venous hemoglobin concentrations and arterial oxygen saturation within the ranges of sea level populations
  - Ethiopia — Beall, ***PNAS***, 2002; Anna Di Rienzo, ***Plos Genetics***, 2007
- High level of hemoglobin concentration or erythrocytosis with arterial hypoxemia
  - Andeans — Bigham, ***Plos Genetics***, 2009
- Increased blood flow/normal venous hemoglobin concentration with or with out arterial hypoxemia
  - Tibetans — Beall. ***PNAS***. 2002, 2007

Table 1. Three patterns of adaptation to high-altitude hypoxia are identified by comparing the presence (+) or absence (–) of erythrocytosis and arterial hypoxemia

	Partial pressure of inspired oxygen, % of sea level	Erythrocytosis	Arterial hypoxemia
Sea level	100	–	–
Ethiopian	64	–	–
Tibetan	60	–	+
Andean	60	+	+

Data were obtained by using the mean values of hemoglobin concentration and oxygen saturation of hemoglobin of sea level populations as a point of reference, published values from Andean and Tibetan high-altitude populations at 4,000 m (12, 17, 18), and the present Ethiopian sample.

Beall. ***PNAS***. 2002

Whether these phenotypic contrasts reflect different genetic adaptations across populations remains an open question.

GWAS studies for Tibetans

- GWAS studies examine the region of the genome that contains the gene ***EPAS1***, which is a transcription factor. Among its functions, ***EPAS1*** helps control the production of red blood cells, increasing their number under low oxygen conditions (hypoxia).

- Some mutations in ***EPAS1*** that increase its expression, are associated with increased hypertension and stroke at low altitude, symptoms similar to that of mountain sickness.

- These facts suggest that Tibetans living permanently at high altitudes might undergo selection at ***EPAS1*** to reduce the fitness consequences of excessive red blood-cell production.

- 31 SNPs were found in intron region of ***EPAS1*** gene which is a transcription factor also called ***HIF2α. EPAS1*** gene were found in high linkage disequilibrium that correlated significantly with hemoglobin concentration in Tibetans population(196 Tibetans and 84 Han individual from HaoMap3, Beall et al. 2010).

- Because all of the found SNPs are located at the **intron region** of ***EPAS1*** gene, the detailed functional association between genotype and phenotype of hypoxia regulatory remains unclear. We still want to know the detailed **type of selections which exists for human high-altitude adaption occurred in the hypoxia-inducible factor (HIF) ?**

Domestic dog

- The dog was the first domesticated animal and has been the most widely kept working, hunting, and pet animal in human history.
- The domestic **dog** (*Canis lupus familiaris*) is a subspecies of the gray wolf (*Canis lupus*), a member of the Canidae family of the mammalian order Carnivora.
- The term “domestic dog” is generally used for both domesticated and feral varieties. The word "dog" can also refer to the male of a canine species, as opposed to the word "bitch" which refers to the female of the species.

- A study of fossil dogs and wolves in Belgium, Ukraine, and Russia tentatively dates domestication from ***14,000*** years ago to more than ***31,700*** years ago. Another recent study has found support for claims of dog domestication between ***14,000*** and ***16,000*** years ago, with a range between ***9,000*** and ***34,000*** years ago, depending on mutation rate assumptions.

- The human settlement history on highland is rather short, which dates from about ***25,000*** years ago (Zhao et al. 2009).

Himalayan mountain dog

(Tibetan Mastiff)






Himalayan mountain dog
(Tibetan Mastiff)

- The **Tibetan Mastiff** is an ancient breed and type of domestic dog (*Canis lupus familiaris*) originating with nomadic cultures of Tibet, China, Nepal, Ladakh, and Central Asia. Tibetan mastiffs live comfortably on the Tibetan Plateau, which with an average elevation of over ***4,500*** meters is nearly ***1,000*** times as high as Shanghai.

- It has been theorized that an early Tibetan dog is the ancestor to all Molossus breeds, although this is disputed by most experts. Some studies found that while most common dog breeds genetically diverged from the wolf approximately ***42,000*** years ago, the Tibetan Mastiff genetically diverged from the wolf approximately ***58,000*** years ago.

Highland settlement: Tibetans Vs. Tibetan Mastiff

- Tibetans settlement history on highland is rather short, which dates from about ***25,000*** years ago.

- A study of fossil dogs and wolves in Belgium, Ukraine, and Russia tentatively dates domestication from ***14,000*** years ago to more than ***31,700*** years ago.

- The **Tibetan Mastiff** genetically diverged from the wolf approximately ***58,000*** years ago.

Tibetans Vs. Tibetan Mastiff

- A reasonable consideration could be: Tibetans and Tibetan Mastiff have the similar settlement history on the Tibetan Plateau. **Tibetan Mastiff** migrated to the plateau with humans(**Tibetans**) about ***25,000*** years ago.

- So, Tibetan mastiffs may ***can be most conveniently used for studying*** ***"What type of selections exists for human high-altitude adaption occurred in the hypoxia-inducible factor (HIF)? "***

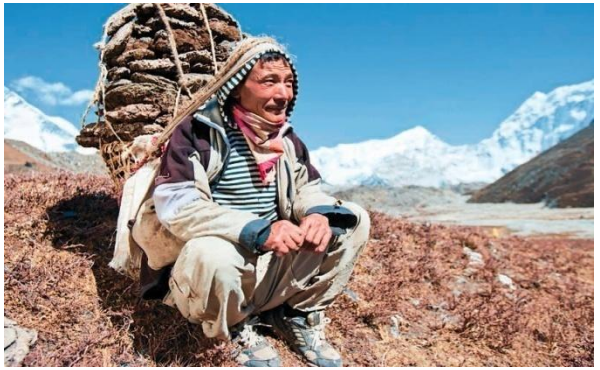
Introduction

Genomes of High-altitude species

- Highland wild animals
  - Yak
  - Tibetan antelope
  - Snow leopard
  - Wild boar
- Highland population
  - Tibetans
  - Andeans
  - Ethiopians

Qiu et al. Nat. Genet. 2012

Li et al. Nat. Genet. 2013



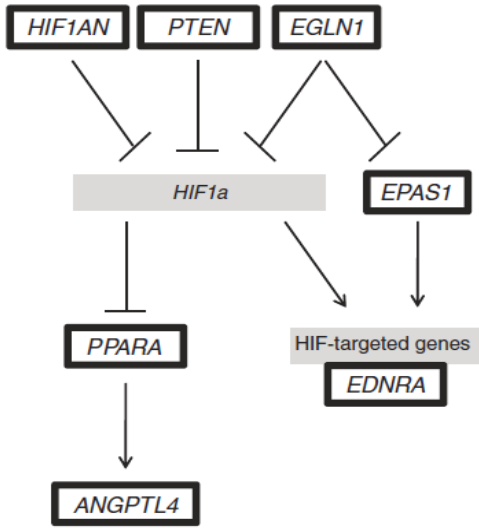
Simonson et al. and Yi et al. Science. 2010

Physiological hallmarks

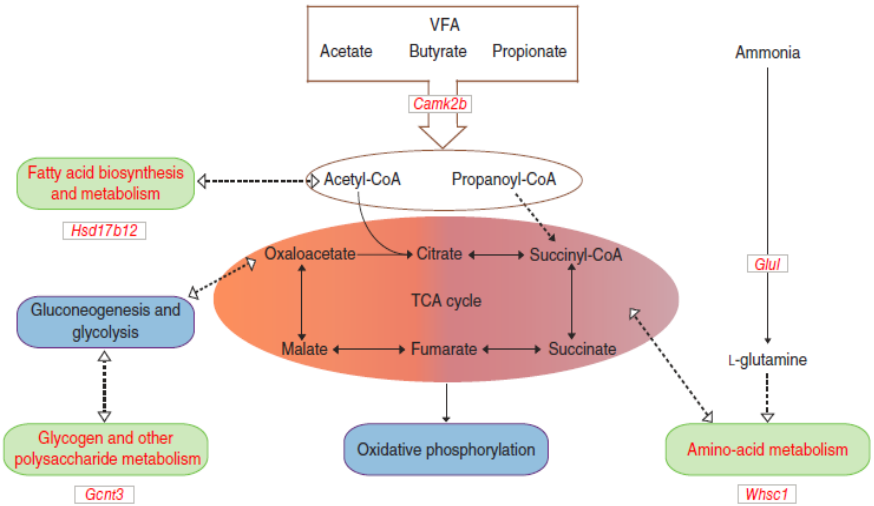
- Little reduction in O2 sat and increase in Hb levels
  - Amhara in Ethiopia
- High level of hemoglobin concentration
  - Andeans
- Increased blood flow
  - Tibetans
- High oxygen affinity of hemoglobin
  - Yak
  - Deer mice
- ?
  - **Tibetan Mastiff**

Pre-genome scan result

- HIF (hypoxia-inducible factor) pathway
  - Tibetans
- Metabolic pathways
  - Yak
  - Tibetan antelope



Simonson et al. Science. 2010



Although a lot of studies focused on wildlife and human highlanders, no research was performed on domesticated animals that migrated to the plateau with humans.

Beall et al. PNAS, 2010
Qiu et al. Nat. Genet. 2012

Results about *EPAS1/HIF2α*

- 31 SNPs were found in intron region of ***EPAS1*** gene which is a transcription factor also called ***HIF2α.*** ***EPAS1*** gene were found in high linkage disequilibrium that correlated significantly with hemoglobin concentration in Tibetans population(196 Tibetans and 84 Han individual from HaoMap3, Beall et al. 2010).

- Because all of the found SNPs are located at the **intron region** of ***EPAS1***, the detailed functional association between genotype and phenotype remains unclear. We still want to know the detailed **type of selections exists for human high-altitude adaption occurred in the hypoxia-inducible factor (HIF)?**

Strategy

Foucs on domesticated animals that migrated to the plateau with humans/Tibecan.

Tibetans **Vs.** *Tibetan Mastiff*

Increased blood flow(Tibetans) **Vs.** *?(Tibetan Mastiff)*

Genome wide association study **Vs.** *Whole genome sequencing*

illumina genotyping chips

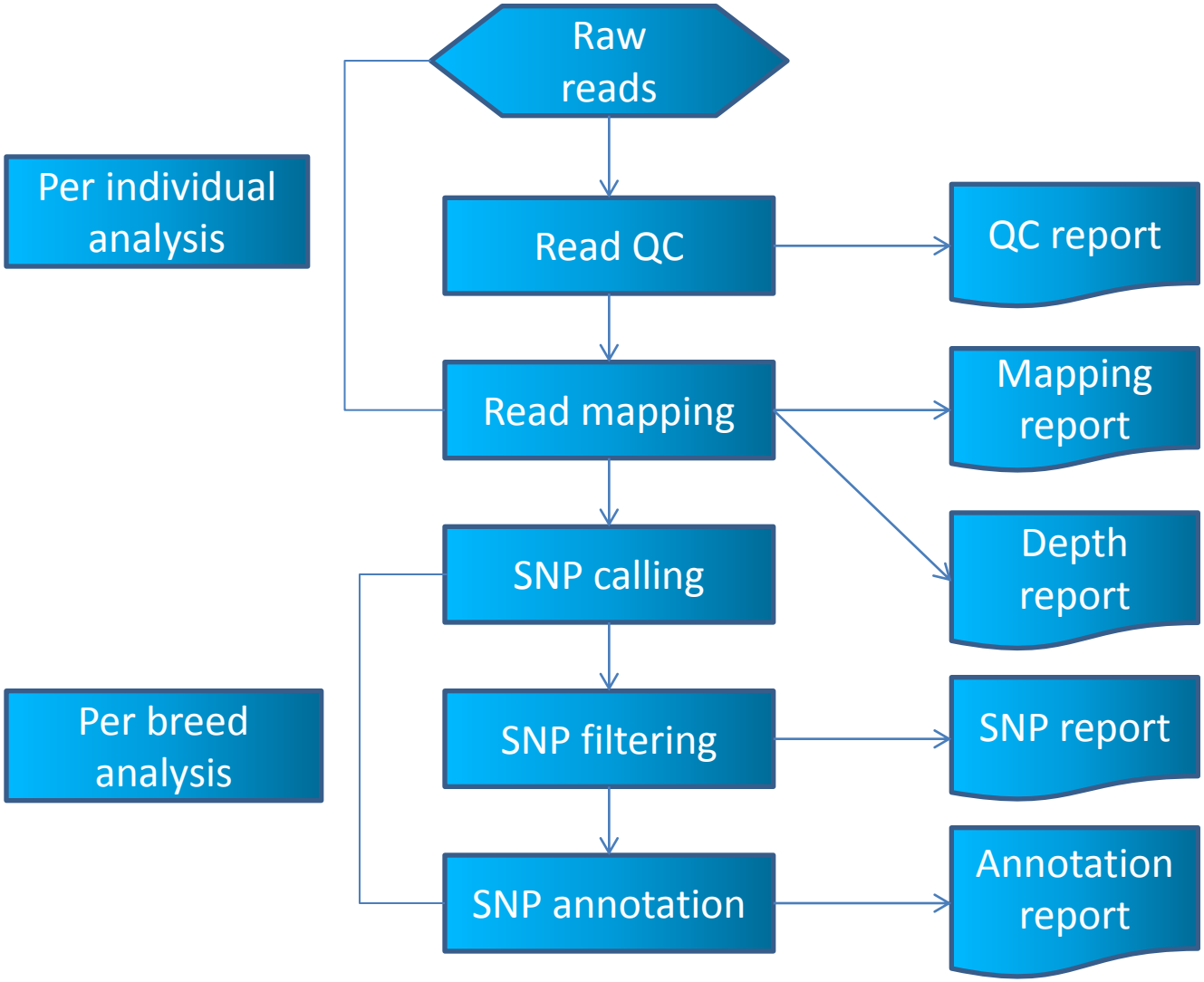
Vs.

Samples and Data

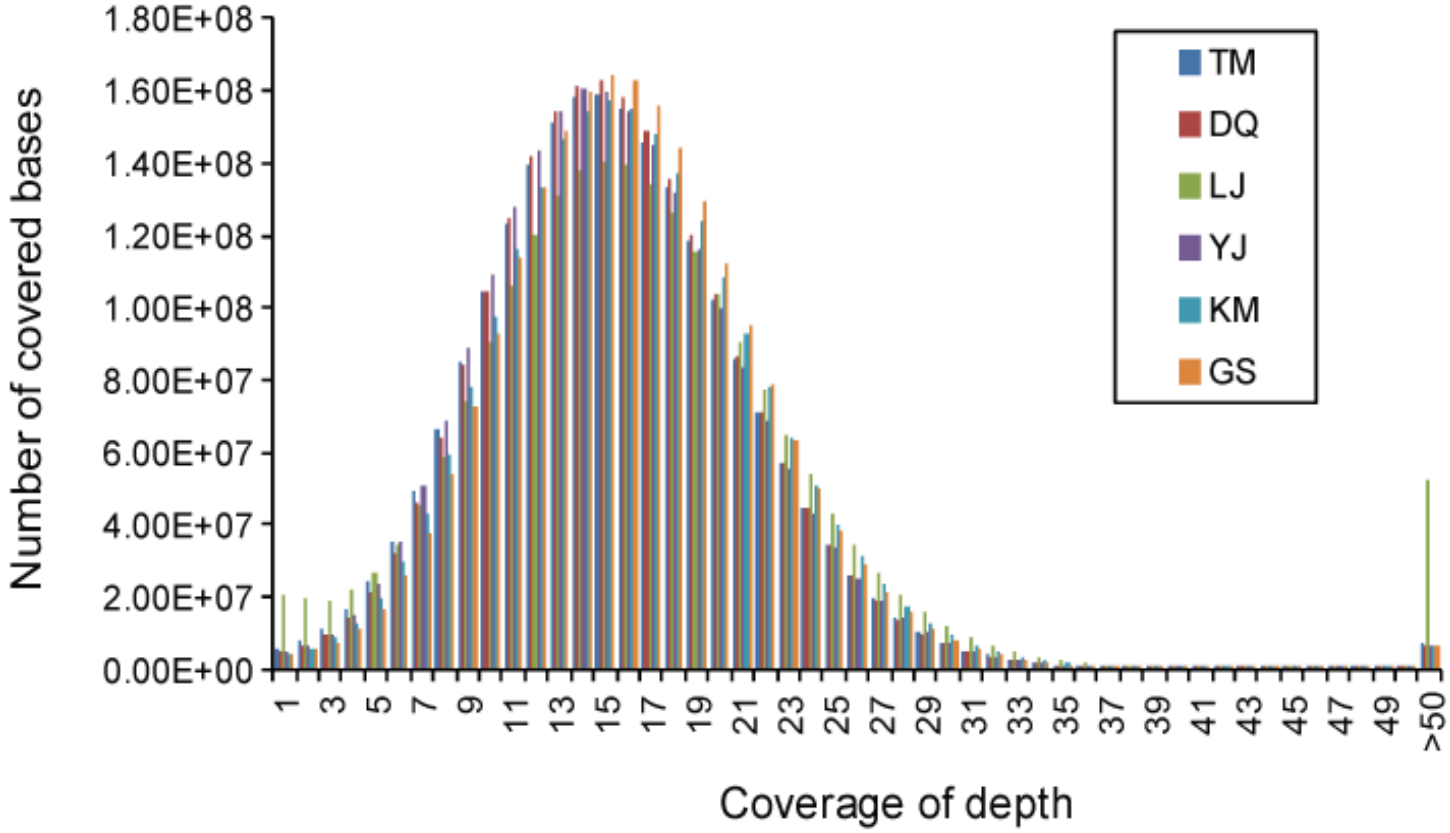
- We sampled six dog breeds from continuous altitudes along the "***Ancient Tea Horse Road***" in southwestern China.
- Each dog was sampled from one individual village to avoid potential kinships.
- The sex ratio was kept as 1:1 for each breed.
- In total, **60** dogs from six dog breeds were sequenced.

Breed (abbreviation)	History	Sample size	Location	Altitude
Tibetan Mastiff (TM)	Ancient	10	Cuomei, Tibet, China (n = 4)	5,100 m
			Yushu, Qinghai, China (n = 4)	4,200 m
			Diqing, Yunnan, China (n= 2)	3,300 m
Diqing indigenous dog (DQ)	Ancient	10	Diqing, Yunnan, China	3,300 m
Lijiang indigenous dog (LJ)	Ancient	10	Lijiang, Yunnan, China	2,400 m
Kunming dog (KM)	Modern	10	Kunming, Yunnan, China	1,800 m
German Shepherd (GS)	Modern	10	Kunming, Yunnan, China	1,800 m
Yingjiang indigenous dog (YJ)	Ancient	10	Yingjiang, Yunnan, China	800 m

From raw reads to SNPs

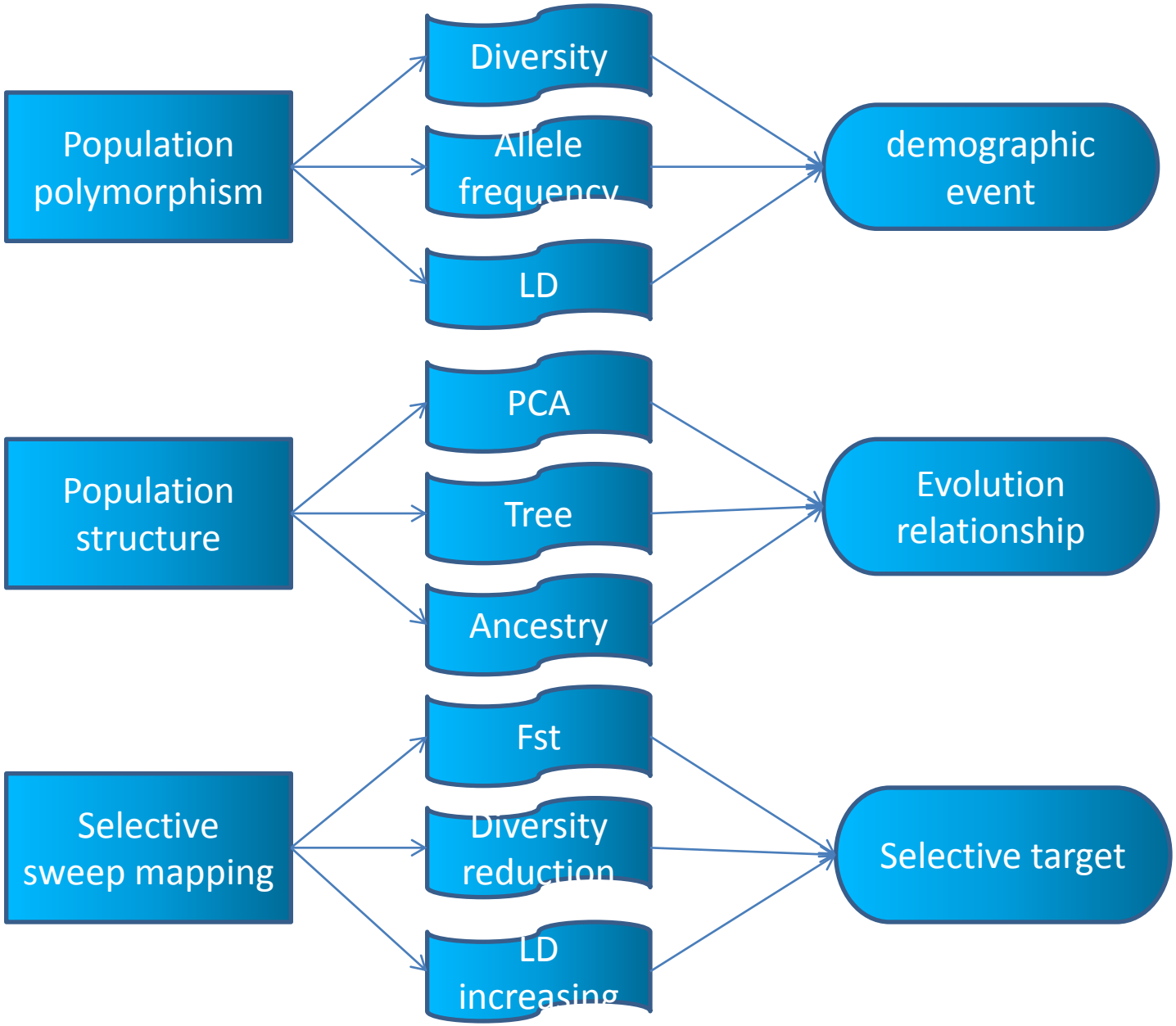


Sequencing depth

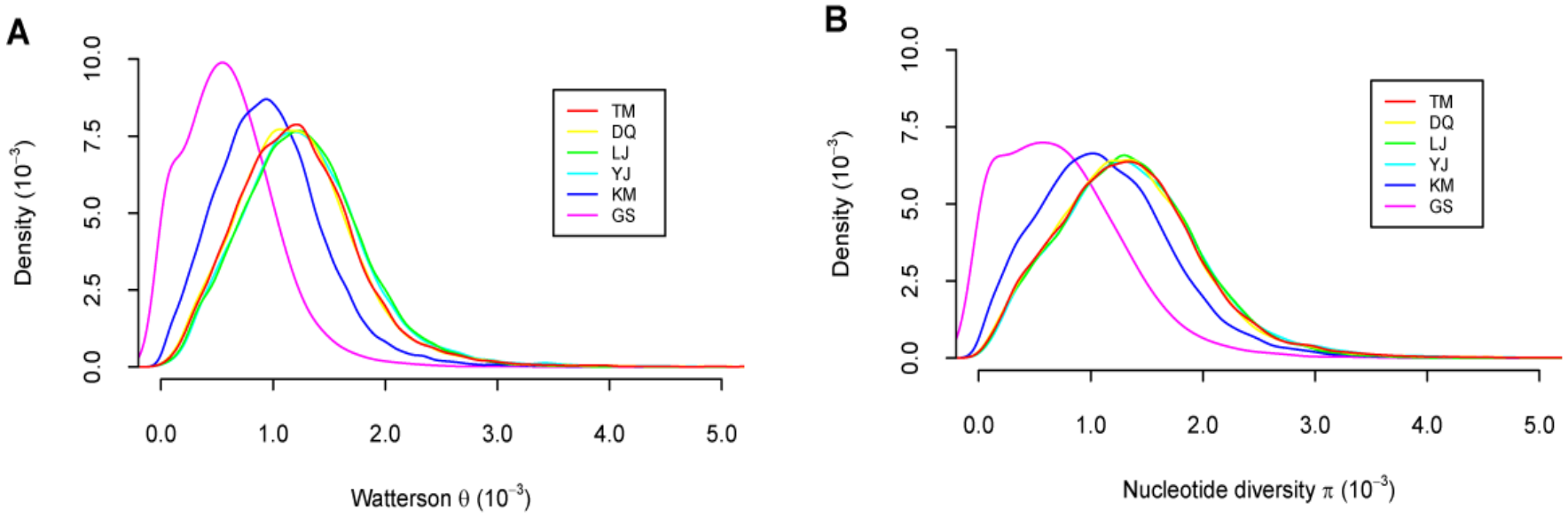


- The effective depth is about **15**× for each dog.

Population genetics analysis

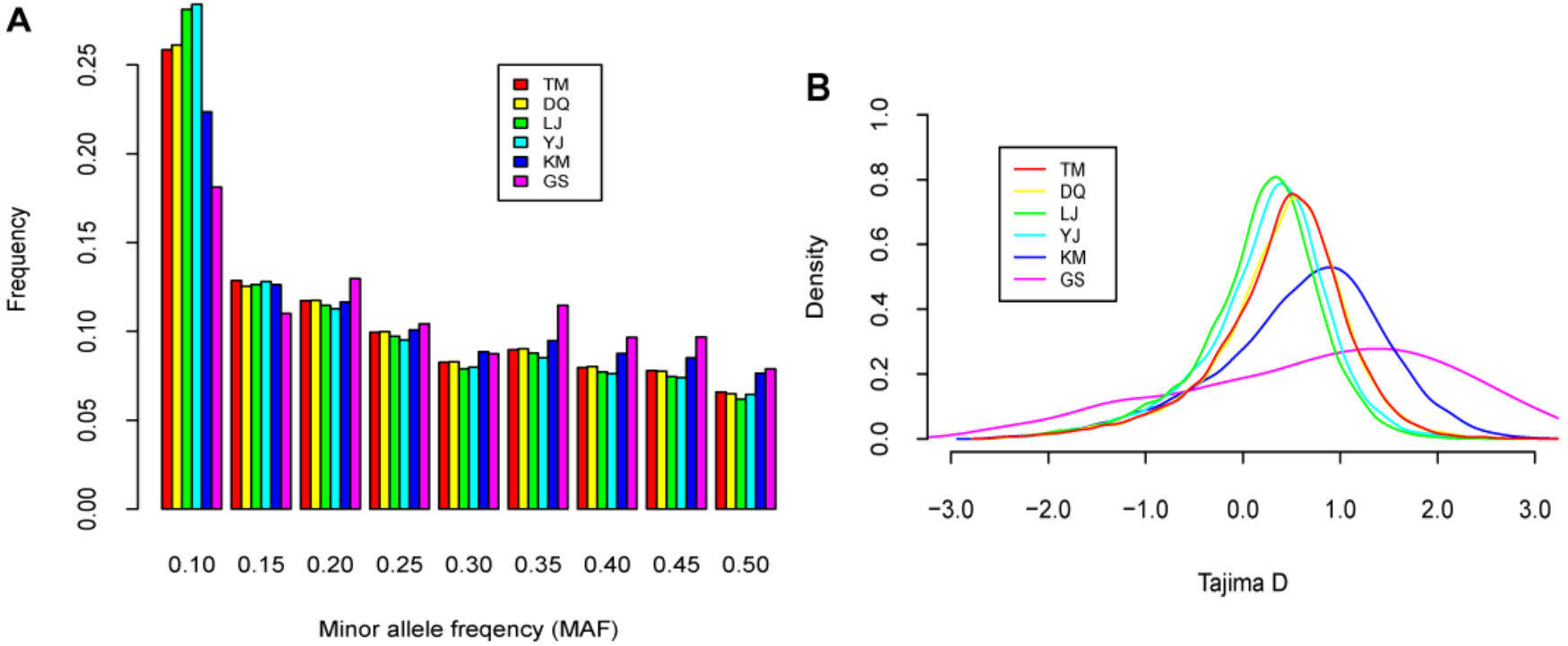

Population polymorphism
Diversity
Allele frequency
LD
demographic event
Population structure
PCA
Tree
Ancestry
Evolution relationship
Selective sweep mapping
Fst
Diversity reduction
LD increasing
Selective target

Genetic diversity



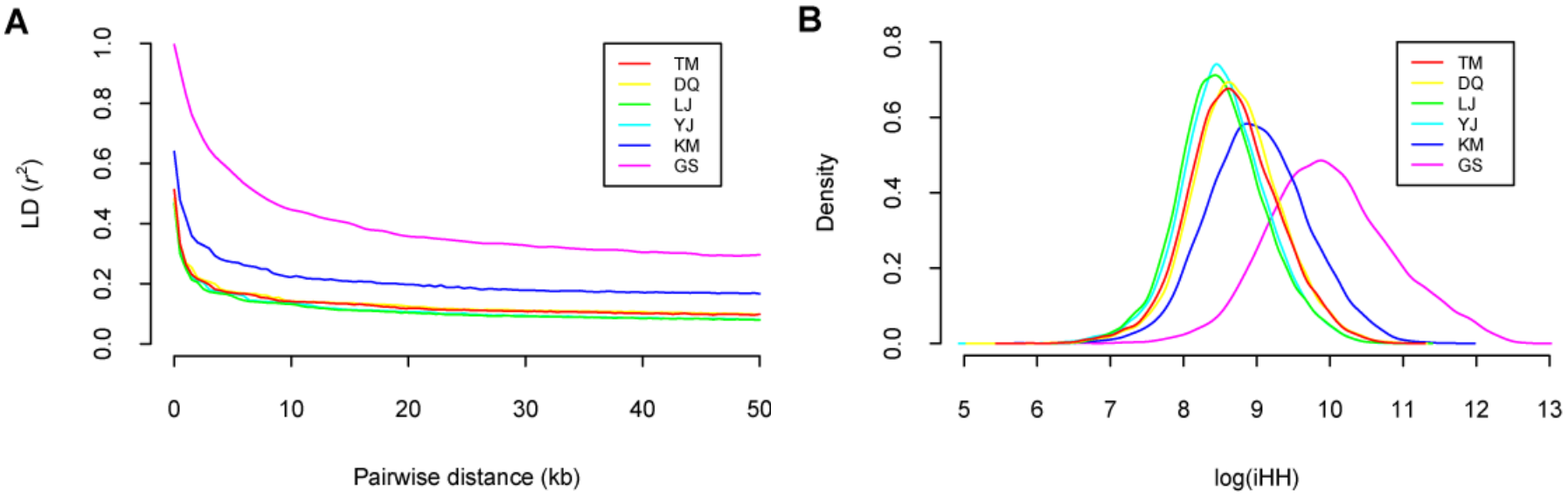
- Genome-wide genetic diversity, as measured by Watterson's ϑ and pairwise nucleotide diversity π, are higher in ancient breeds (TM, DQ, LJ and YJ) than those in modern breeds (KM, GS).

Allele frequency spectrum



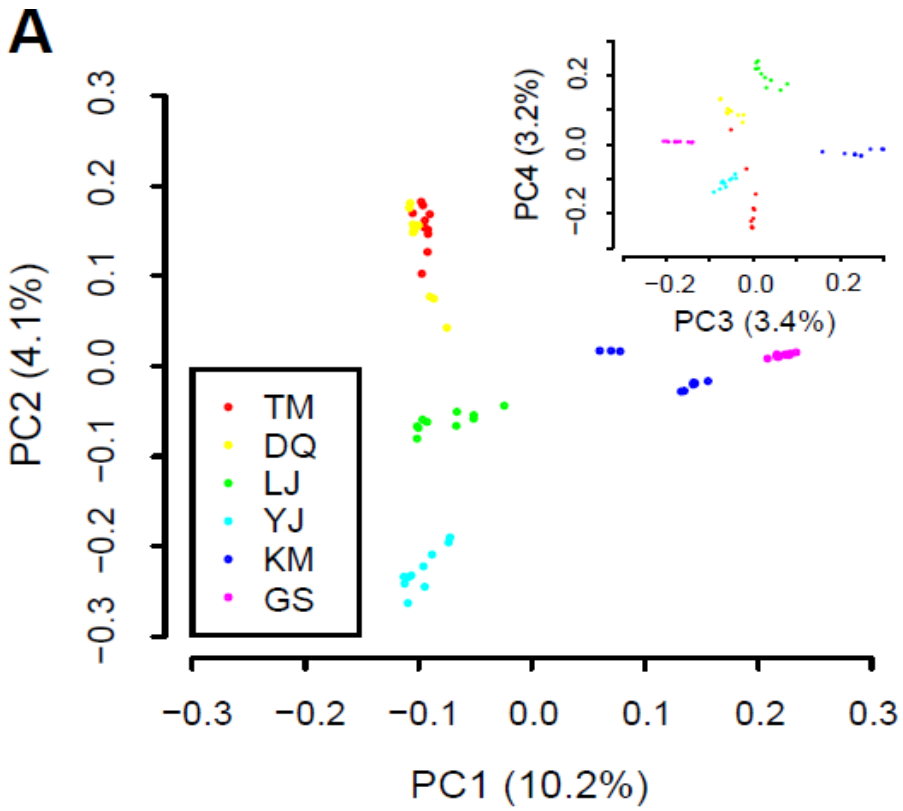
- Minor allele frequency (MAF) distribution and Tajima's *D* indicate a larger proportion of low frequency minor alleles in the ancient breeds than in the modern breeds.

Linkage disequibrium (LD)



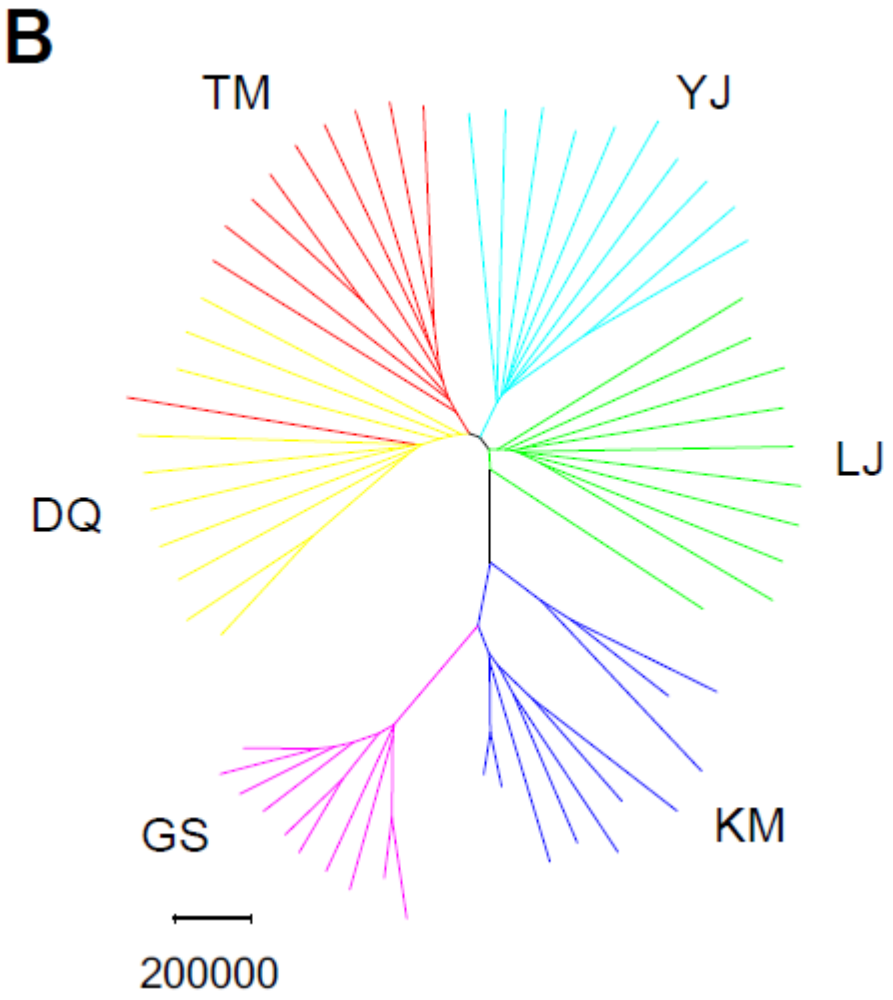
- Ancient breeds have a faster decay of pairwise correlation coefficient (r^2) and lower integrated haplotype homozygosity (iHH) than the modern breeds.

PCA clustering



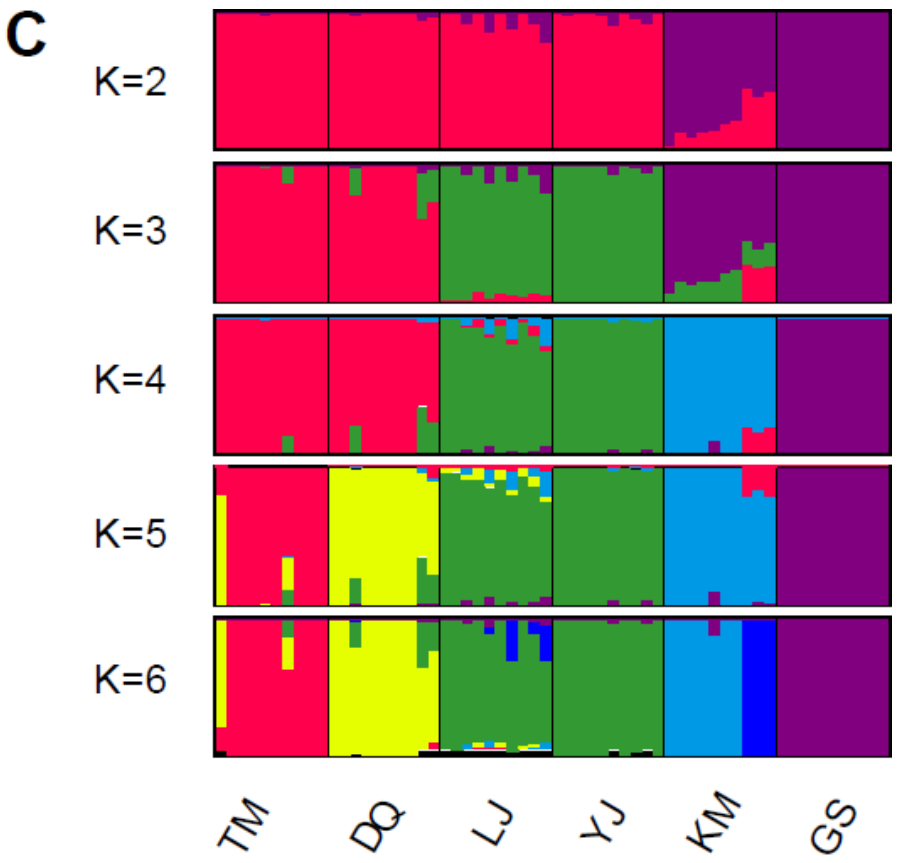
- LJ, YJ, KM and GS could be separated by the first and second eigenvectors, while TM and DQ are mixed.
- However, when more eigenvectors are incorporated, most dogs in TM and DQ could still be separated.

Phylogenetic tree



- TM and DQ are close but split into different branches (except one dog in TM).

Ancestry inference



Ancient and modern division

High and middle/low altitude division

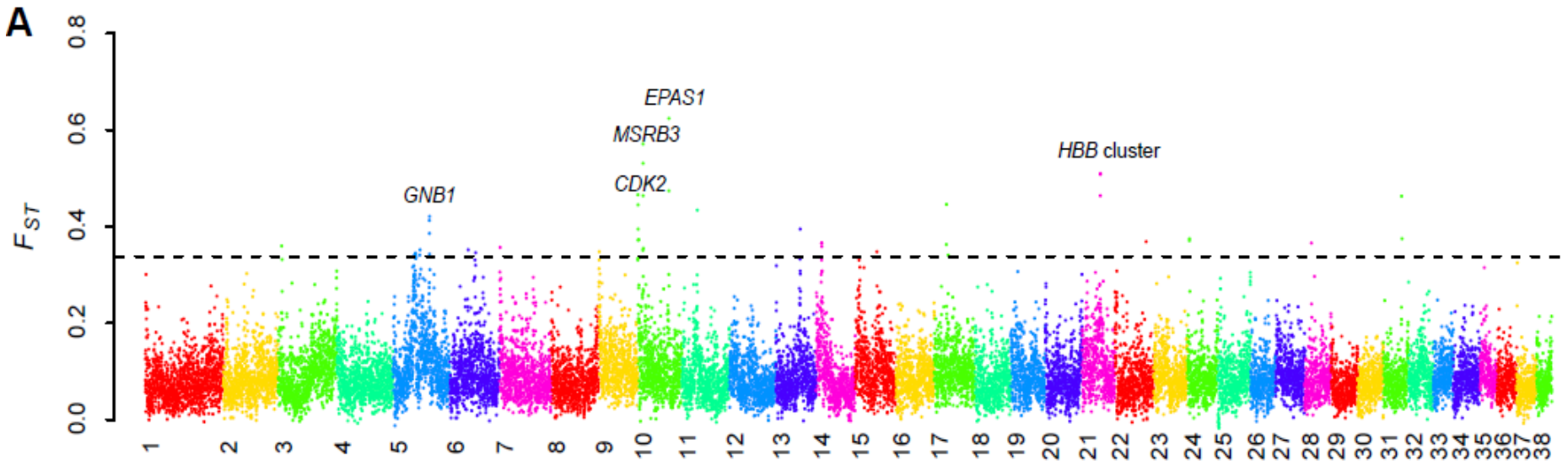
TM and DQ division

- TM and DQ have close but distinct genetic background.
- Slight extent of ingression happened from DQ to TM.

Comparison among altitudes

- We classified the breeds into three altitude levels:
  - high (>3,000 m): TM, DQ
  - middle (2,000-3,000 m): LJ
  - low (<2,000 m): YJ, KM, GS
- Grouping multiple breeds living at similar altitudes could contribute to reducing the influence of breed-specific genetic background.

Whole-genome F_{ST} mapping



- We performed whole-genome F_{ST} scan and focused on regions with the extreme F_{ST} value ($Z(F_{ST}) > 5$) .
- 28 unique autosomal regions containing 141 candidate genes were identified.
- Five genes of them including: *EPAS1*, *MSRB3*, *HBB*, *CDK2* and *GNB1* belong to the GO categories 'response to oxygen levels' and 'response to oxidative stress'.

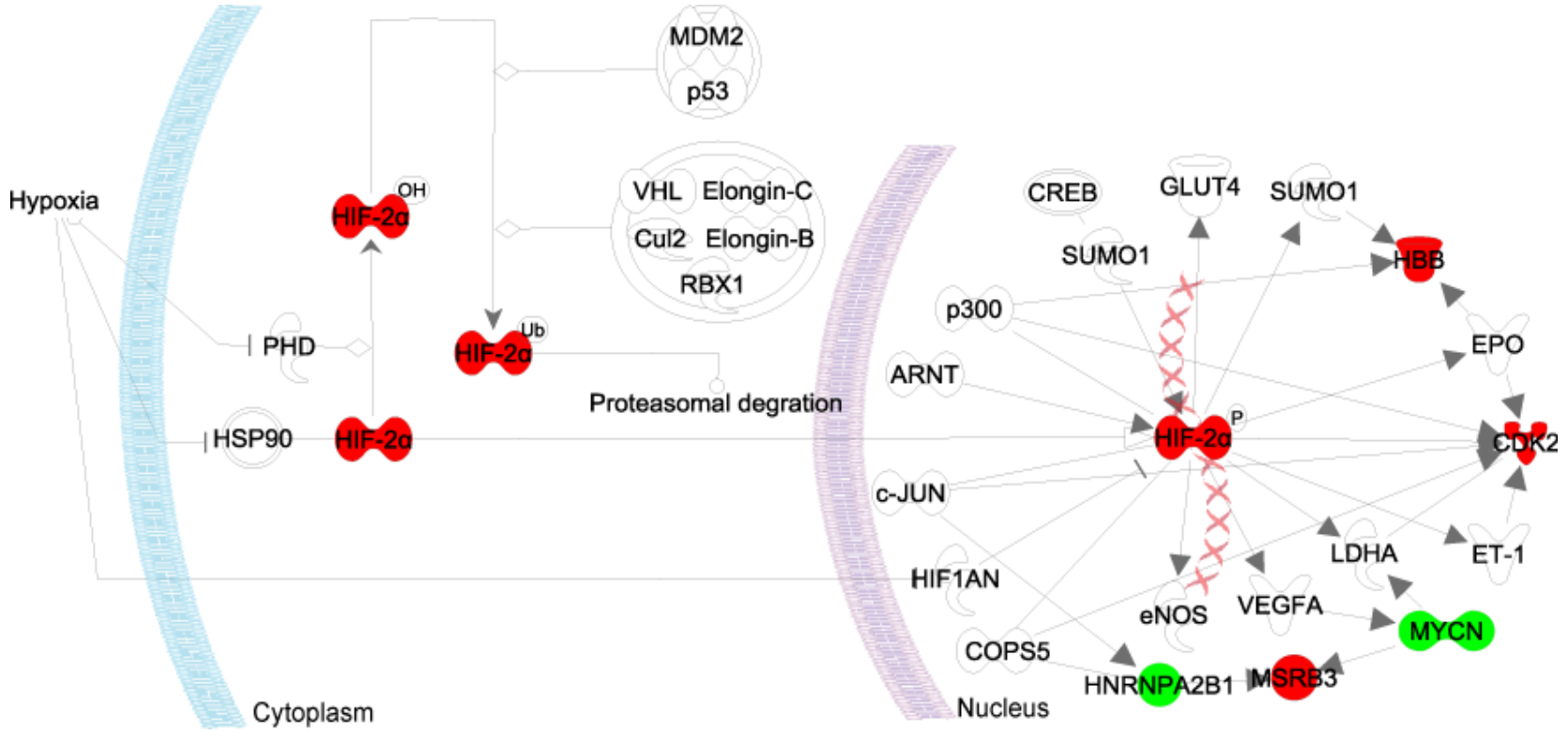
F_{st}: genetic differences among population

Results

Five oxygen-response genes

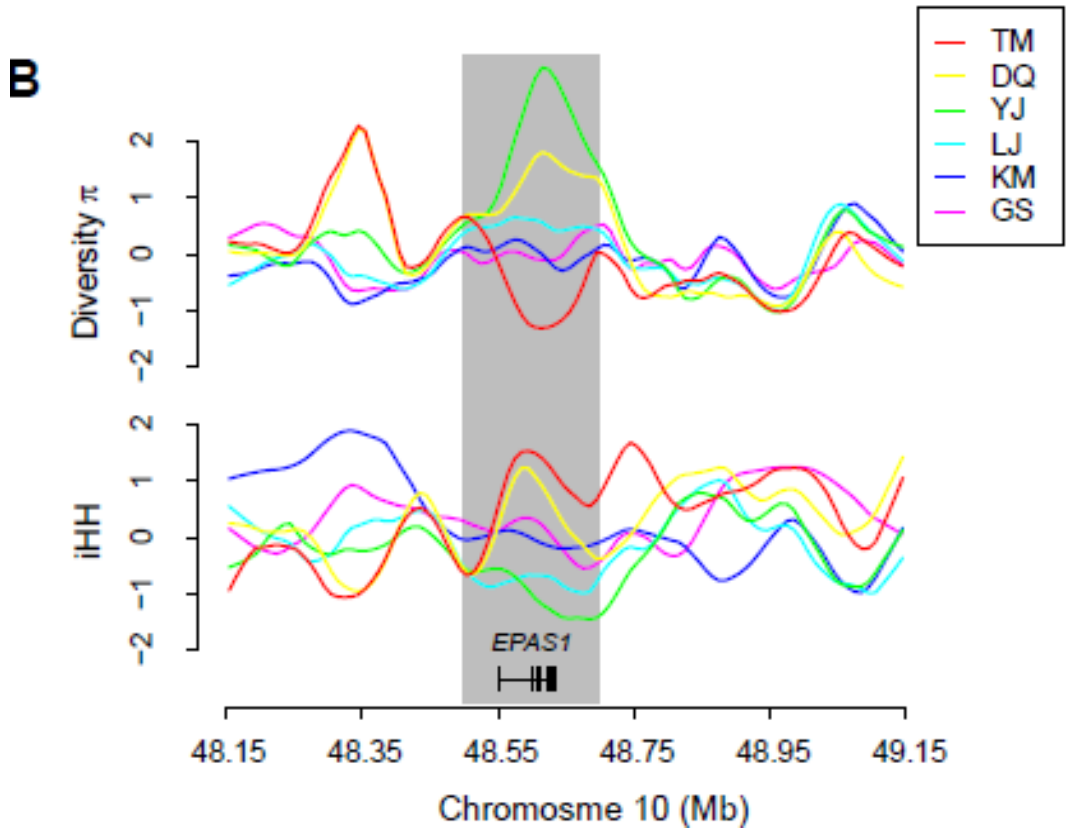
Sweep region	Candidate gene	$Z(F_{ST})$				Sweep date
		Overall	Middle vs low	High vs middle	High vs low	
Chr10: 48600000-48700000	*EPAS1*	10.830	1.844	8.024	10.968	Middle to high
Chr10: 8000000-8100000	*MSRB3*	9.975	6.749	1.861	9.767	Low to middle
Chr21: 28100000-28200000	*HBB*	8.507	4.142	3.466	8.838	Low to high
Chr10:300000-400000	*CDK2*	7.182	4.973	4.592	6.927	Low to high
Chr5: 56700000-56800000	*GNB1*	6.693	5.804	3.595	5.474	Low to middle

HIF pathways



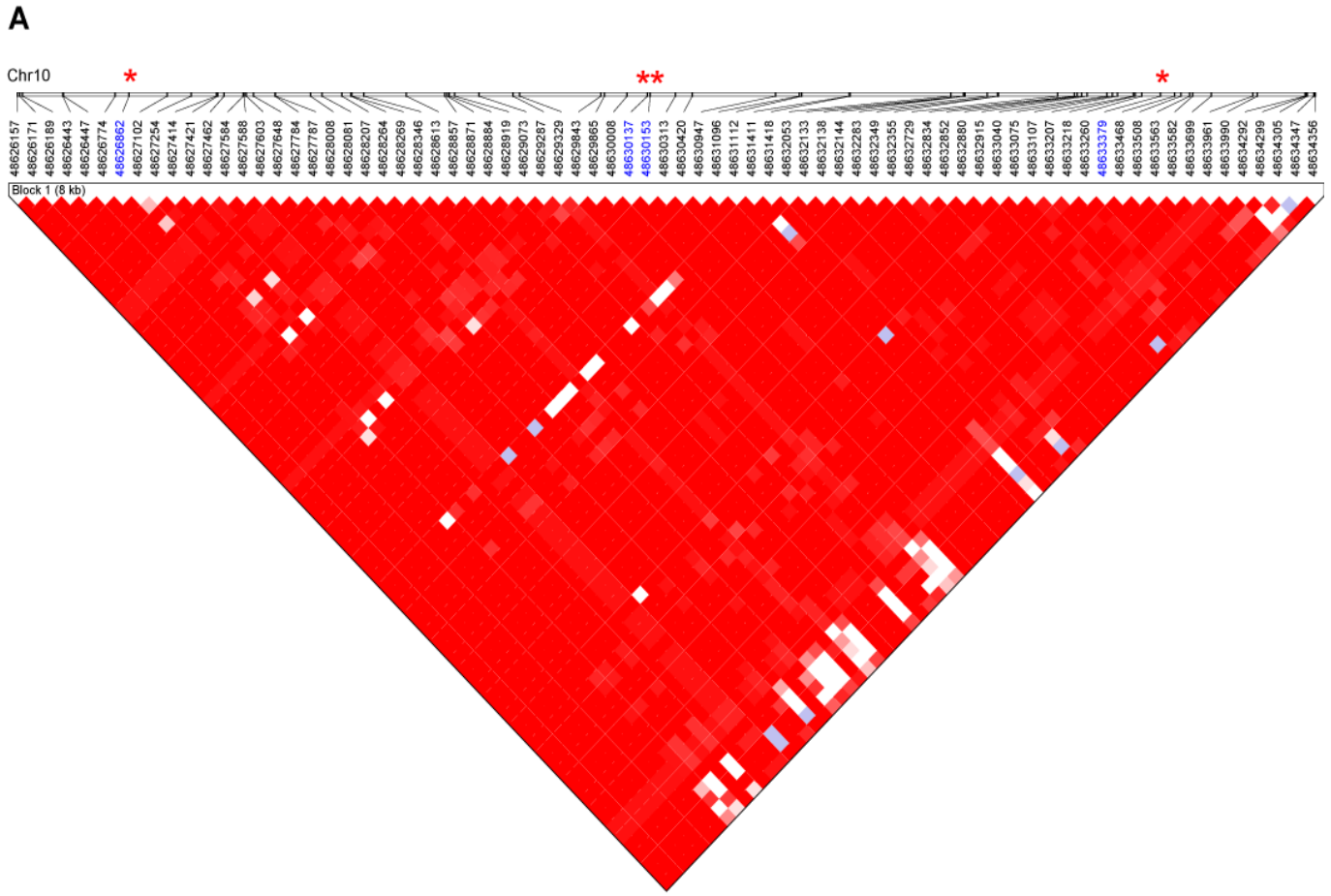
- The region with the strongest differentiation *EPAS1*, a gene encodes the hypoxia-inducible factor (HIF) 2A.
- Network analysis indicated that the other candidate hypoxia-response genes we identified would all be regulated by HIF signaling pathway, suggesting an essential role of ***EPAS1*** in the adaption of high-altitude dogs.
- Interestingly, *EPAS1* was also identified as a selective target in Tibetan people.

Other evidence for *EPAS1* sweep



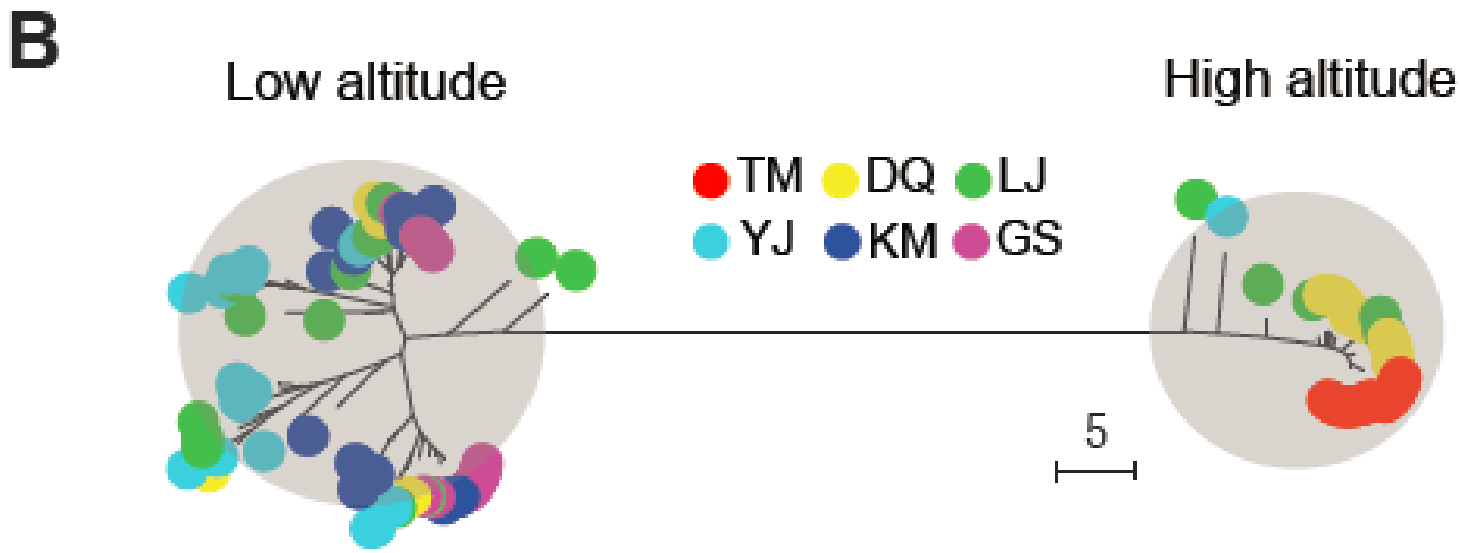
- TM shows a decrease in π and an increase in iHH, supporting the EPAS1 sweep.
- DQ shows an increased in iHH value but an increase in π, which may be caused by genetic admixture.
- LJ has the highest diversity, suggesting extensive admixture in the middle- altitude dogs.

Mutations in *EPAS1*



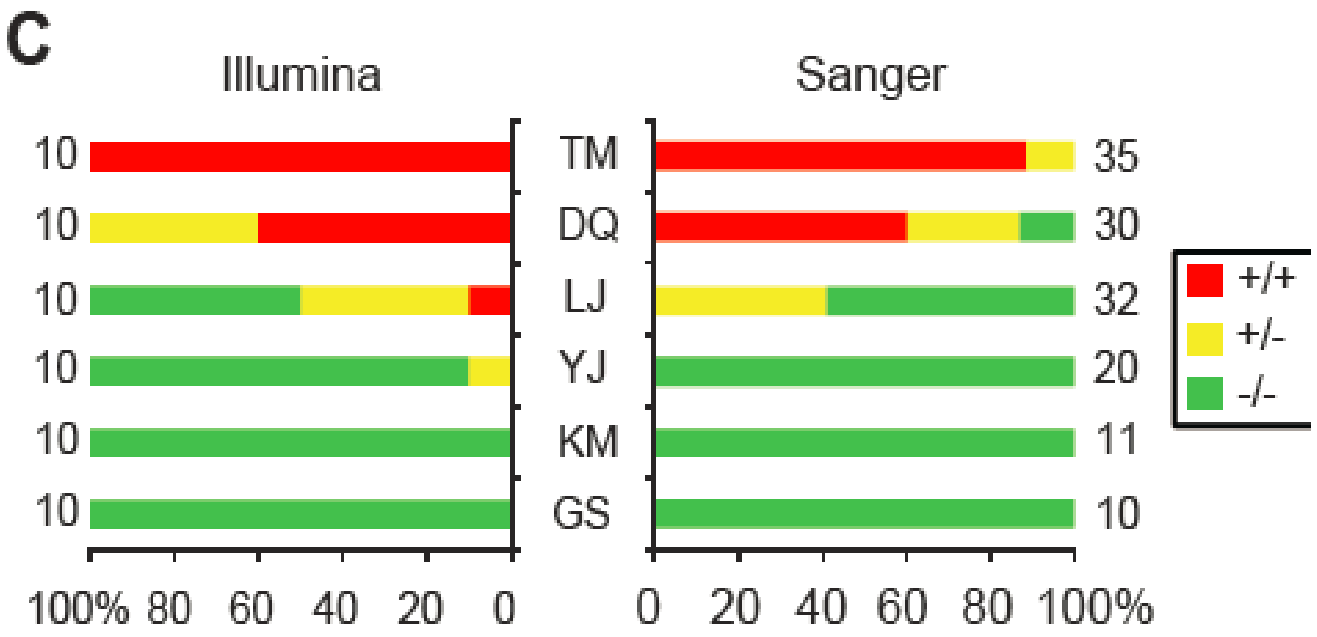
- Four non-synonymous mutations are identified.
- They are in complete linkage disequilibrium (pairwise $r^2 = 1$) in all breeds, which are part of a 8-kb LD block.

Haplotype tree



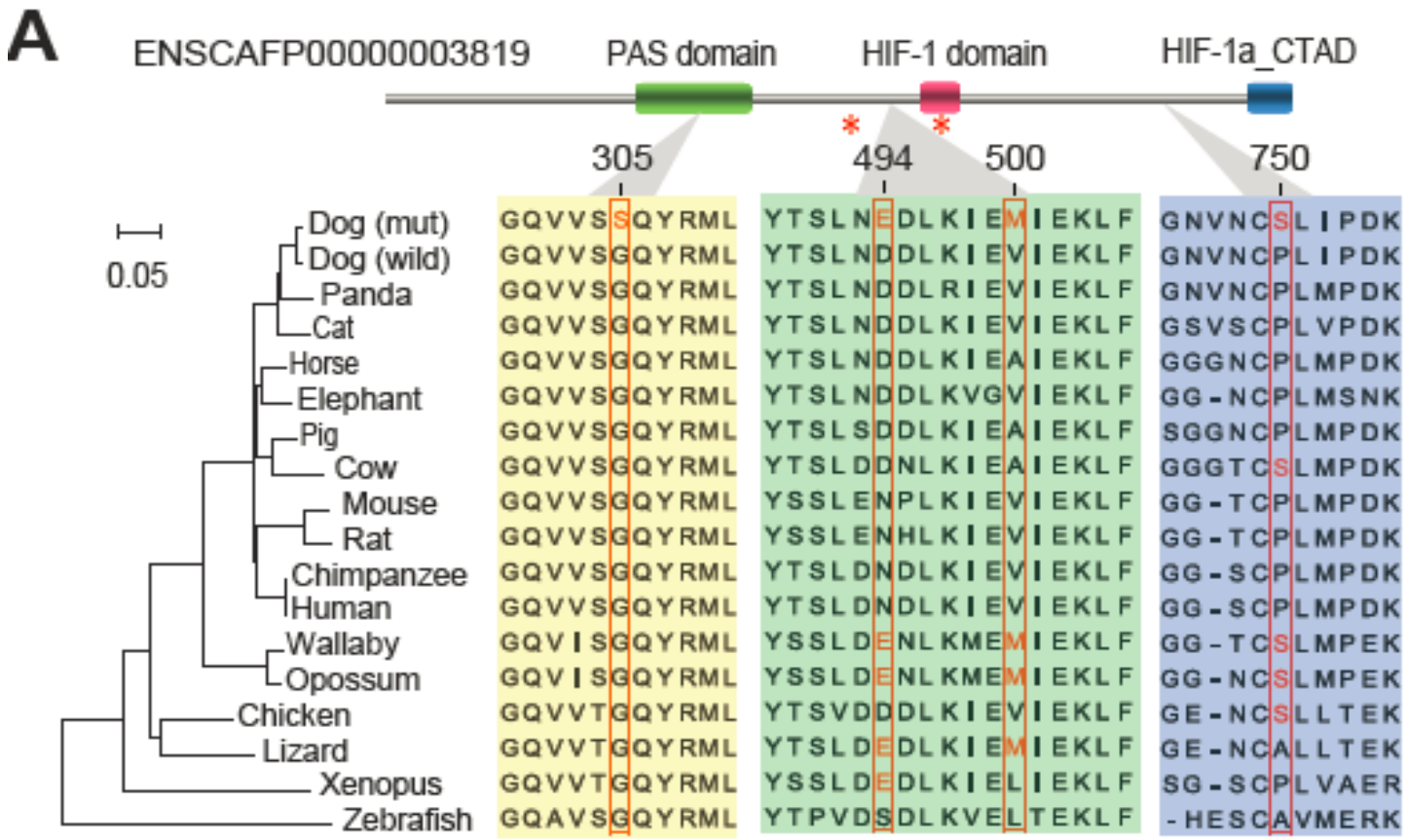
- Haplotypes comprising the four mutant alleles all belong to the high-altitude clade, which have a long evolutionary distance from the haplotypes of the low-altitude clade.

Allele frequency and altitude



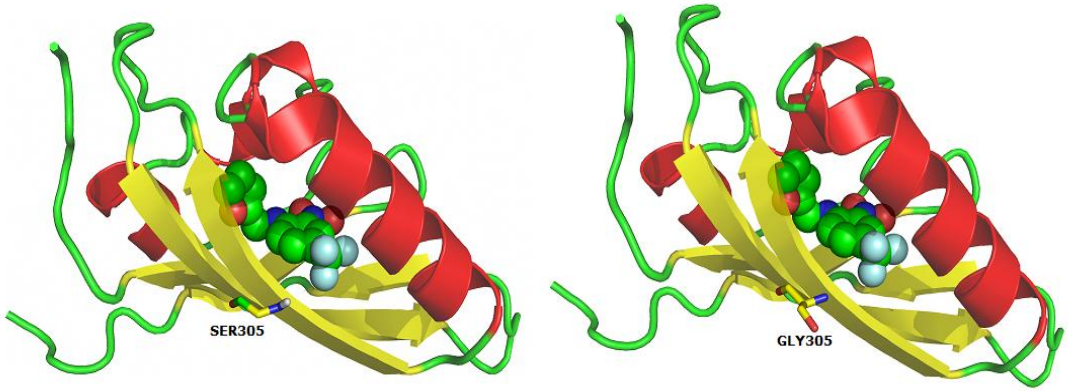
- The frequency of the four mutant alleles (+)is increased with the elevated altitude.
- We performed genotyping of the four alleles on additional samples with Sanger sequencing and the result was similar.

Amino acid conservation



- Among the four non-synonymous mutations, one (**G305S**) occurred in the PAS domain, which is essential for the activity of EPAS1.
- **G305S** is also a quite conserved amino acid mutation, which is invariant among all the vertebrates we examined.

Structural and functional effects of G305S



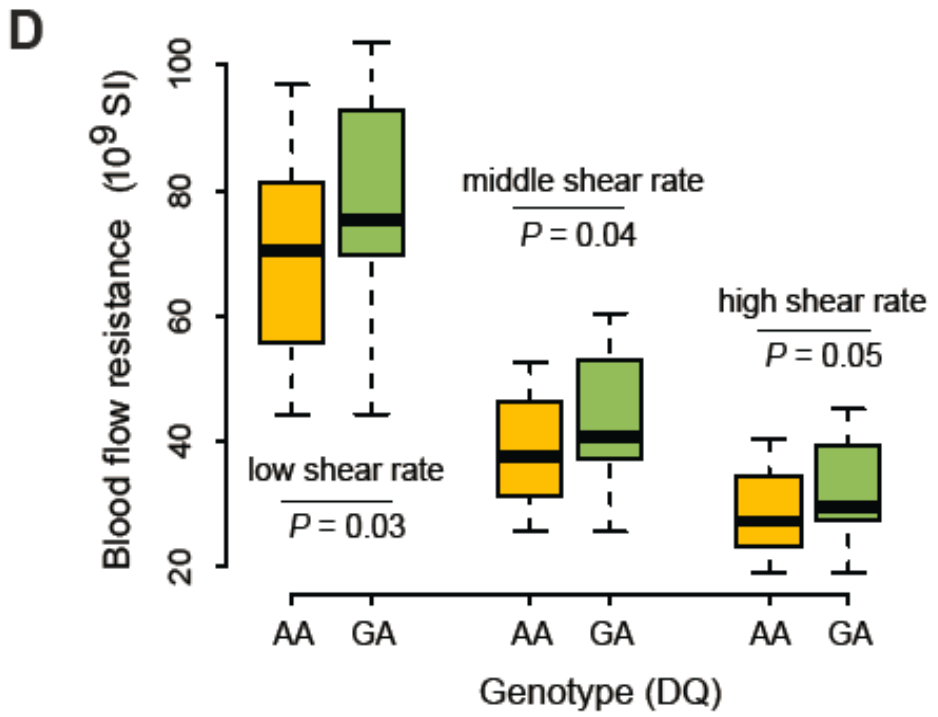
E-value: **2.07e-75**, bit-score: **244**, aligned-length: **113**, Identity to query: **97%**

```
                10        20        30        40        50        60        70        80
           ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
EPAS1  238 PLDSKTFLSRHSMDMKFTYCDDRITELVGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGYV 317
3H82_A   2 GLDSKTFLSRHSMDMKFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGYV  81

                90       100       110
           ....*....|....*....|....*....|...
EPAS1  318 WLETQGTVIYNPRNLQPQCIMCVNYVLSEIEKN 350
3H82_A  82 WLETQGTVIYNPRNLQPQCIMCVNYVLSEIEKN 114
```

- **G305S** occurred in a beta sheet, which could affect the thermodynamic stability of the domain.
- Prediction of functional effects supports that only **G305S** is deleterious, while the other three are tolerated.

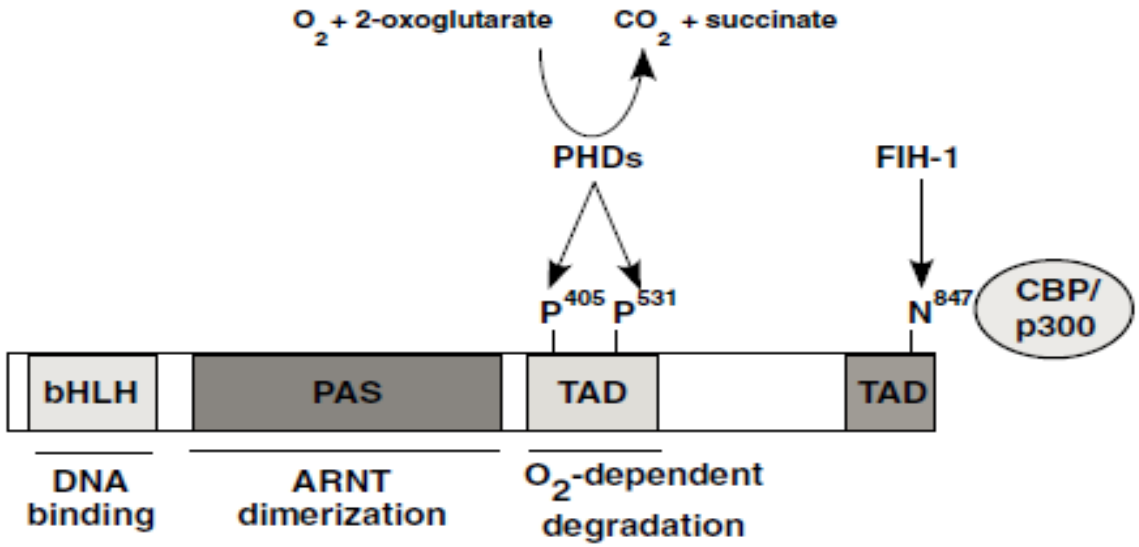
Physiological association



- We conducted association testing for the variant **G305S** and hematologic parameters in DQ, the high-altitude breed where enough homozygotes (*n* = 40) and heterozygotes (*n* = 29) could be collected.
- Although no evident relationship with hemoglobin concentration was found, The homozygotes with two mutant alleles (AA) show decreased vascular resistance as compared with the heterozygotes (GA).

Discussion

Biochemistry of EPAS1

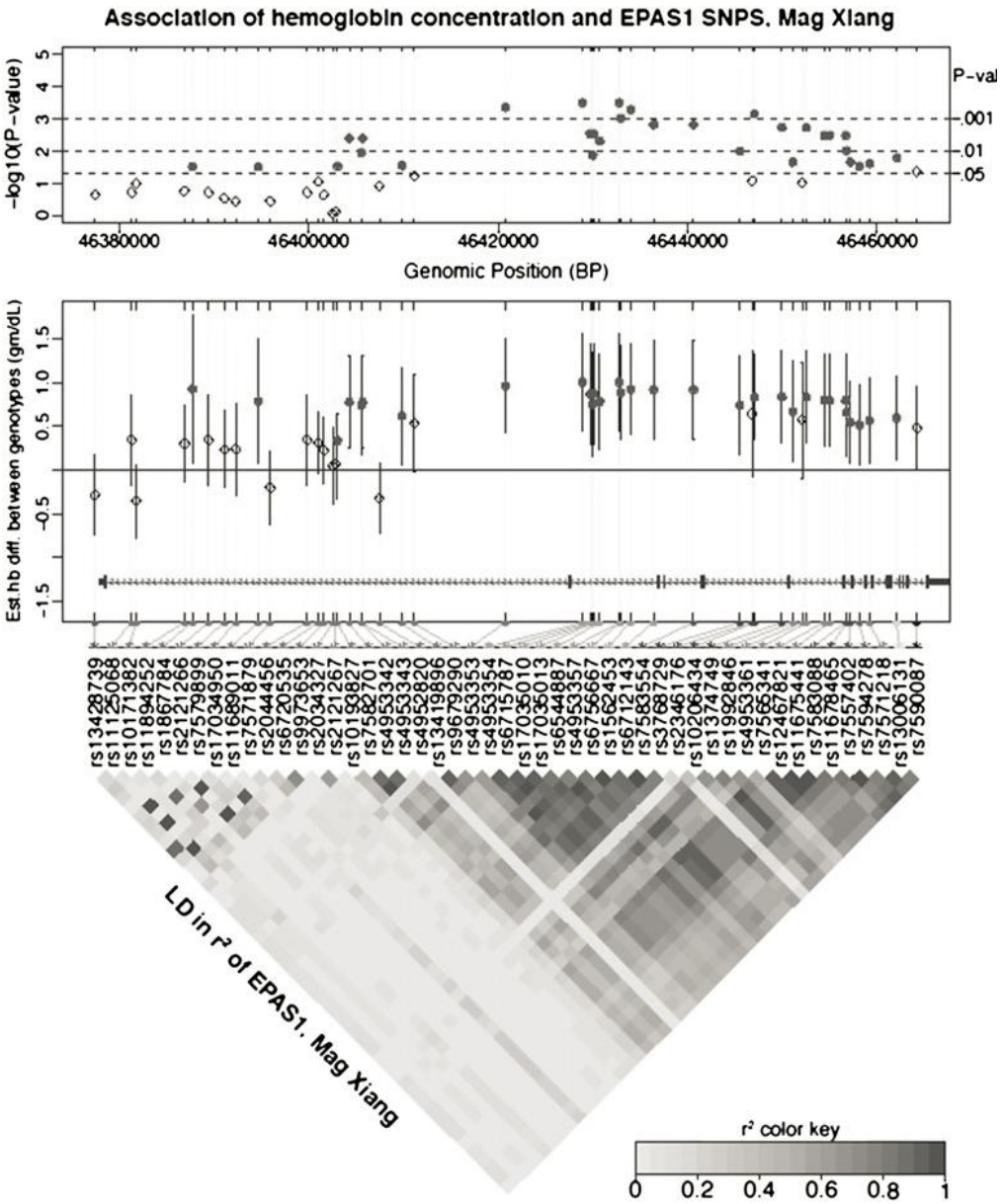


Patel and Simon. *Cell Death Differ.* 2008

Figure 1 Structure of HIF-2α, including the bHLH DNA-binding domain, PAS domain, and transactivation domain (TAD). Prolyl hydroxylases (PHDs) hydroxylate proline residues 405 and 531 in the oxygen-dependent degradation (ODD) of HIF-2α under normoxic conditions, targeting it for degradation by the proteosome. In addition, hydroxylation of asparagine 847 in the C-terminal TAD by factor-inhibiting HIF (FIH-1) inhibits interaction with the co-activators CBP/p300

- The transcription factor encoded by ***EPAS1*** plays a key role in transcriptional response to hypoxia in various physiological and pathologic conditions, including up-regulating the expression of erythropoietin (*EPO*).

EPAS1 and hematologic phenotypes



- Clinical case studies found that several ***gain-of-function*** mutations occurred in close vicinity to the primary hydroxylation site (Pro531) at EPAS1 cause erythrocytosis.
- ***EPAS1*** polymorphisms unique to the native Tibetan people were found to be associated with their lower hemoglobin concentrations, suggesting a ***loss-of-function*** role of ***EPAS1*** in high-altitude adaption.

Beall et al. PNAS. 2010

- **As erythrocytosis is a common symptom of chronic mountain sickness which will lead to high blood viscosity and cardiovascular disorders, the decrease in hemoglobin level may provide a protective mechanism for Tibetan people.**

- In fact, comparing with Andeans and other highland migrants, Tibetans do not show markedly elevated hemoglobin concentration but show a higher blood flow for oxygen delivery, probably due to their unique ***EPAS1*** genotypes.

- However, all ***EPAS1*** variations detected previously in the Tibetan population are at introns, inhibiting further study of its molecular mechanism in the adaptive process. Our studies may shed new insights into the study of the ***EPAS1*** gene function.

Parallel evolution of humans and dogs

- The same selective target, ***EPAS1***, was identified in both Tibetan people and Tibetan Mastiffs.
- The key amino acid mutation in dogs, **G305S**, is predicted to be damaging, which is also likely to cause the loss of function of ***EPAS1***.
- The variation is associated with lower blood flow resistance in high-altitude dogs, which could help to improve hemorheologic fitness.
- Therefore, we propose that similar mechanisms may be adopted by both Tibetans and TM dogs in adaption to high-altitude hypoxia.

Summary

- We performed whole-genome sequencing of 60 dogs from continuous altitudes along the "***Ancient Tea Horse Road*** **"** in the Tibetan Plateau with a sequencing depth of 15× for each dog, which is the largest set of whole-genome sequenced dogs to date.
- It is the first study to characterize the genetic polymorphisms of Tibetan Mastiffs, which are known as "oriental deiform dog" in Tibet and one of the most ancient and ferocious dogs in the world.
- Tibetan Mastiffs have a close but different genetic background with other indigenous dogs living at high altitude. The existence of admixture raises the demands for genetic testing in the identification and breeding of Tibetan Mastiffs.
- We identified several candidate genomic regions with high extent of differentiation among different altitudes. Two of the genes, *EPAS1* and *HBB*, were previously reported as targets of natural selection for hypoxia adaption in humans and other animals.
- We discovered and validated four novel non-synonymous mutations at *EPAS1* in high-altitude dogs, of which G405S occurred in a well defined domain and quite conserved site. This provides a new opportunity to study the role of *EPAS1* in the adaptive processes.

Whole genome sequencing of six dog breeds from continuous altitudes reveals adaption to high-altitude hypoxia

Xiao Gou, Zhen Wang, Ning Li, et al.

Genome Res. published online April 10, 2014
Access the most recent version at doi:10.1101/gr.171876.113

Acknowledgement

- Shanghai Institutes for Biological Sciences
  - Dr. Guohui Ding
- College of Animal Science and Technology, Yunnan Agricultural University
  - Dr. Guo Xiao
- Encode Genomics Biotechnology Co., Ltd.
  - Dr. Feng Qiu
  - Ze Xu

Thanks !