BOTANICKÝ ÚSTAV AV ČR v.v.i.

esf european social fund in the czech republic | EUROPEAN UNION | MINISTRY OF EDUCATION, YOUTH AND SPORTS | OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

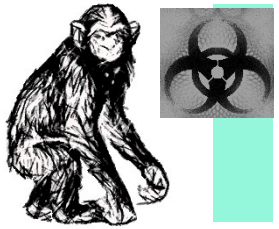
INSTITUTE OF VERTEBRATE BIOLOGY ACADEMY OF SCIENCES CR

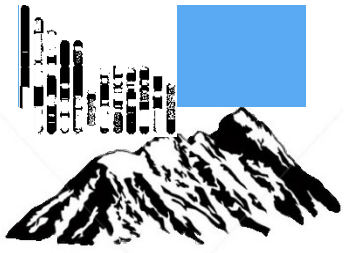
GENETIC SIGNATURES OF NATURAL SELECTION

Jamie Winternitz

Institute of Botany and Vertebrate Biology, Czech Academy of Sciences

Outline of talk





1. **The Chimp and the River**
   - Negative-frequency dependent selection
   - Phylogenetic methods
2. **The Island Fox**
   - Balancing selection
   - Accounting for demography
3. **Men in the Mountains**
   - Positive selection
   - Genome scans

A strange set of symptoms

- 1980s USA
- Opportunistic infections
- Ubiquotious fungus Pneumocystis jirovecii
- Oral candidiasis (yeast)
- Depleted wbc counts (thymus-dependent lymphocytes)
- Kaposi's sarcoma

Something is wrong with the immune system

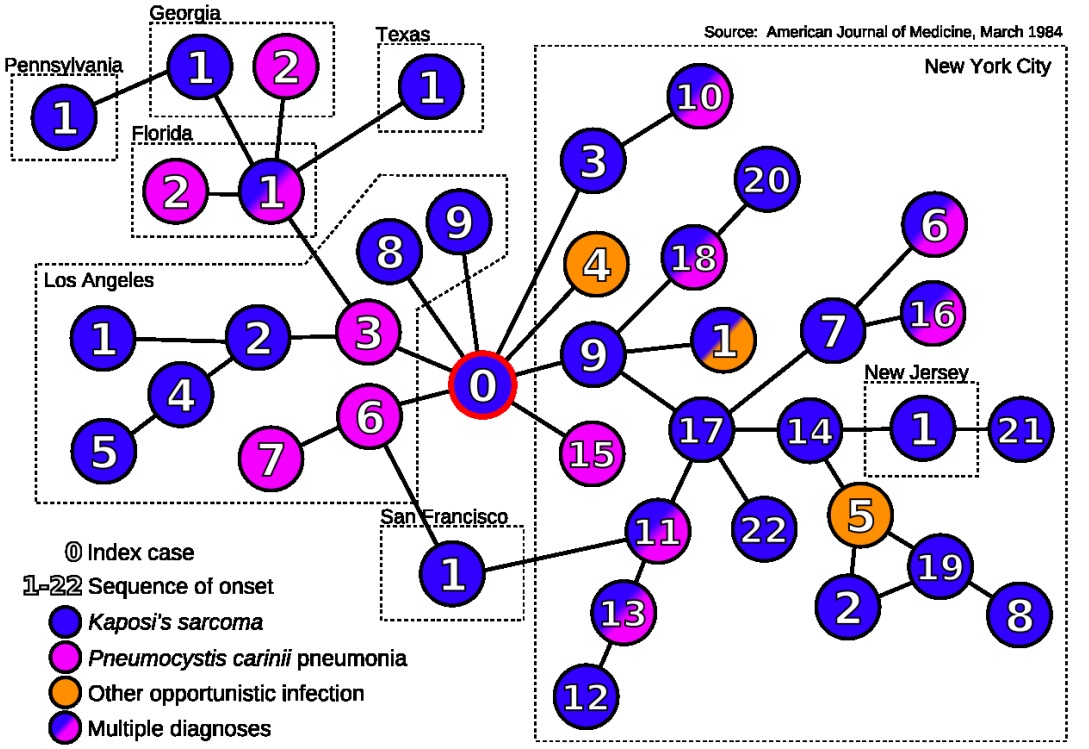
Clusters of infection

- AIDS high incidence in homosexuals linked by sexual interactions -> infectious disease
- Incidence among intravenous drug users -> blood-borne
- Cases among hemophiliacs who received processed/filtered blood transfusions ->must be a virus

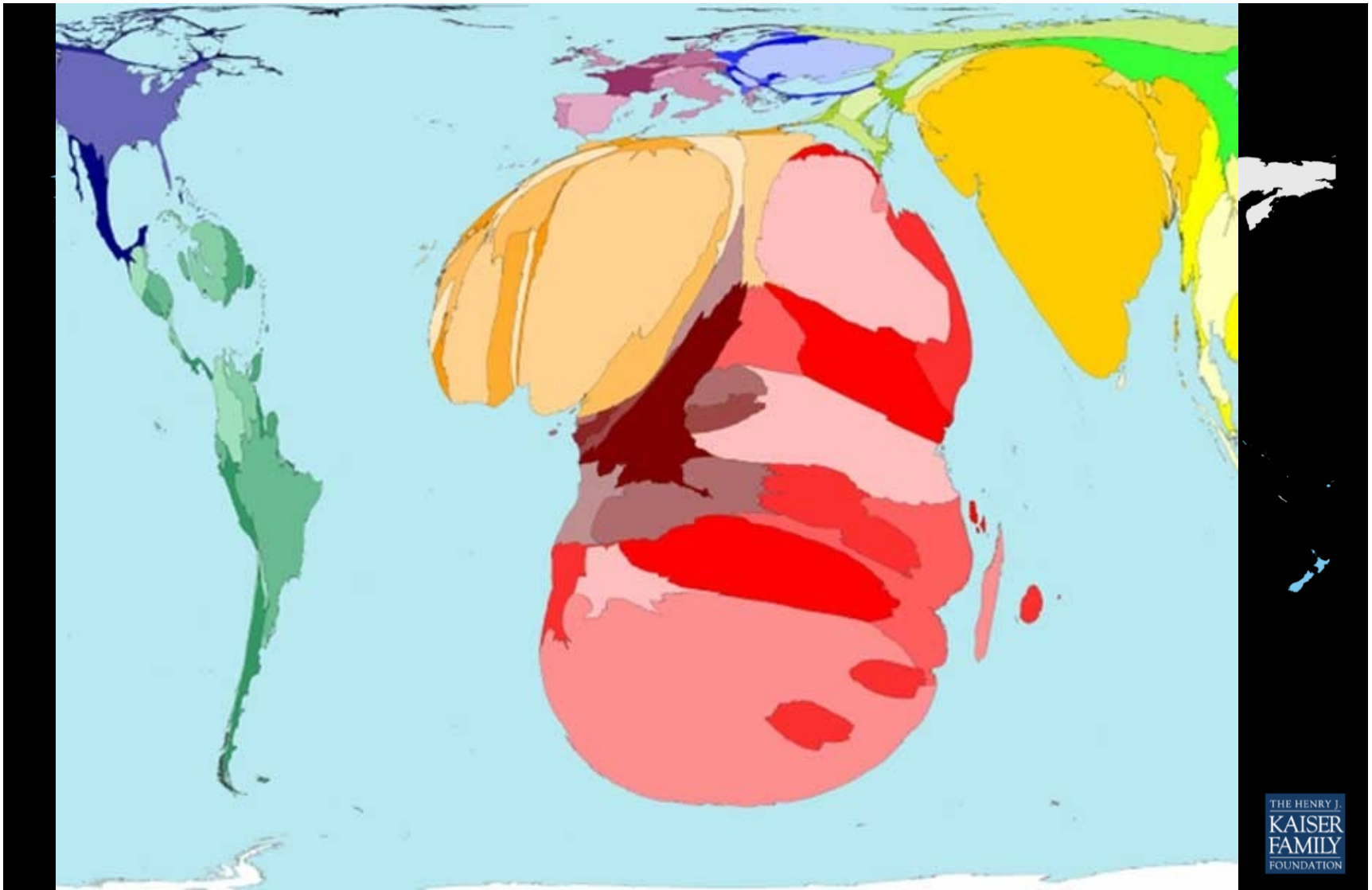
“Patient 0” (Zero)

- A Canadian airline steward named **Gaëtan Dugas** was referred to as "Patient 0" in an early AIDS study by Dr. William Darrow of the CDC

2500 sexual partners

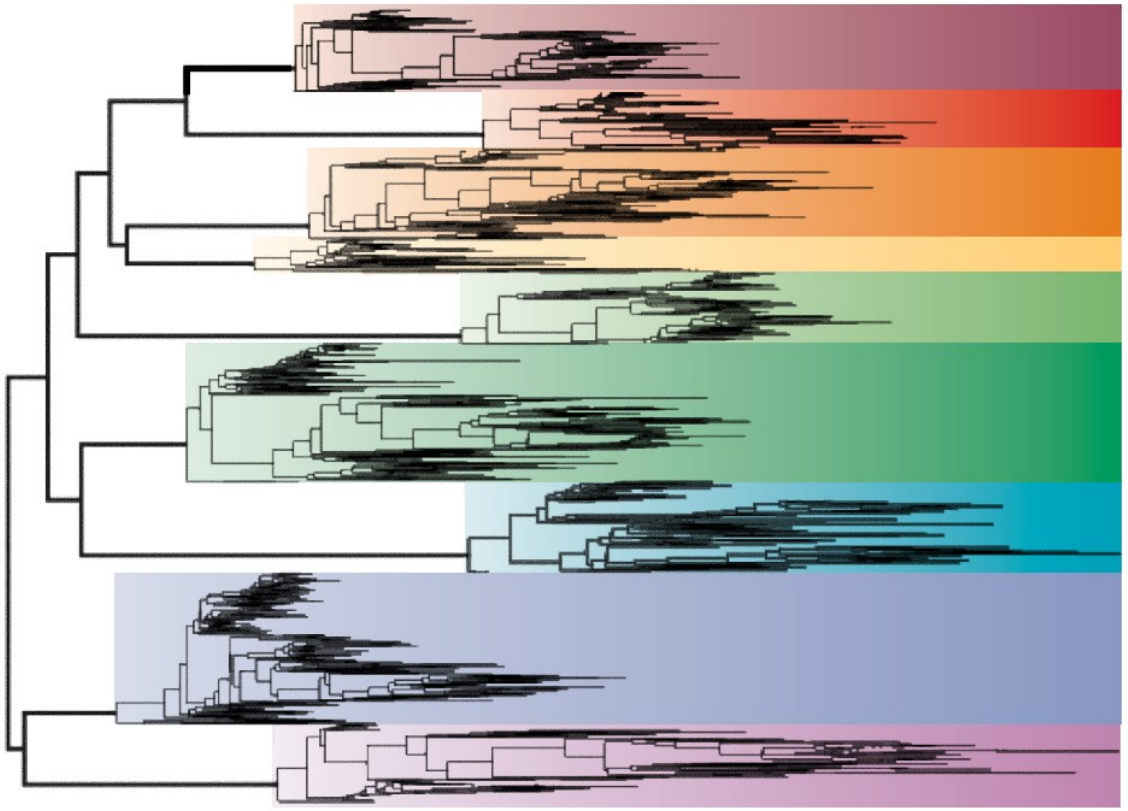


HIV Worldwide



HIV variation

- Retrovirus (Reverse transcription)
- No proofreading = high error rate
- For a virus with a genome about 10 thousand bases in length, that means that basically every time HIV replicates itself, it makes a mistake.
- High viral production 10^8 copies per day
- Recombination, genetic drift, genetic shift, bottlenecks and immune-driven selection



HIV Types & subtypes

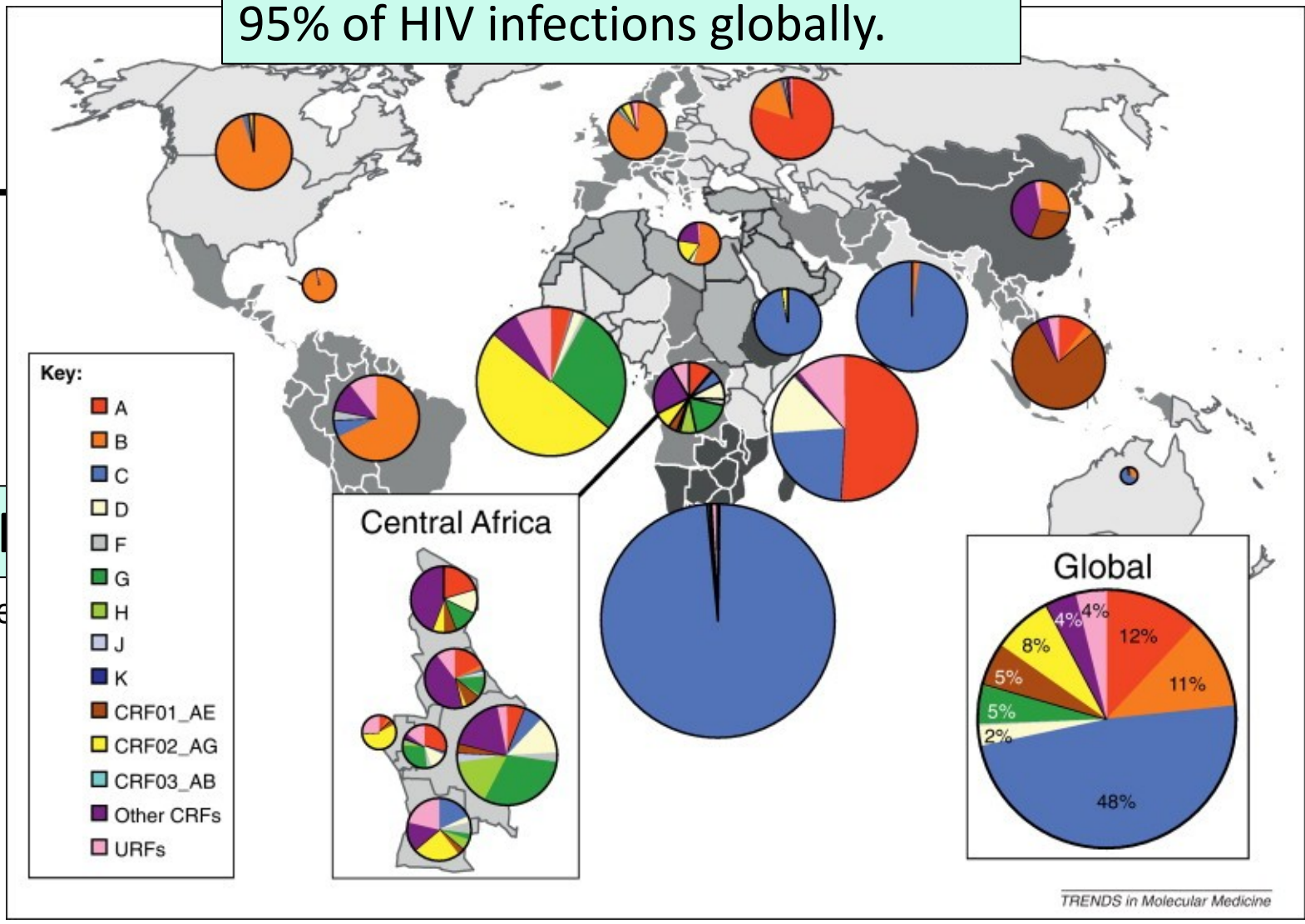
HIV-1 group M is responsible for 95% of HIV infections globally.

HIV-1

Group M

A B C D

Worldwide



SIV in captive primates

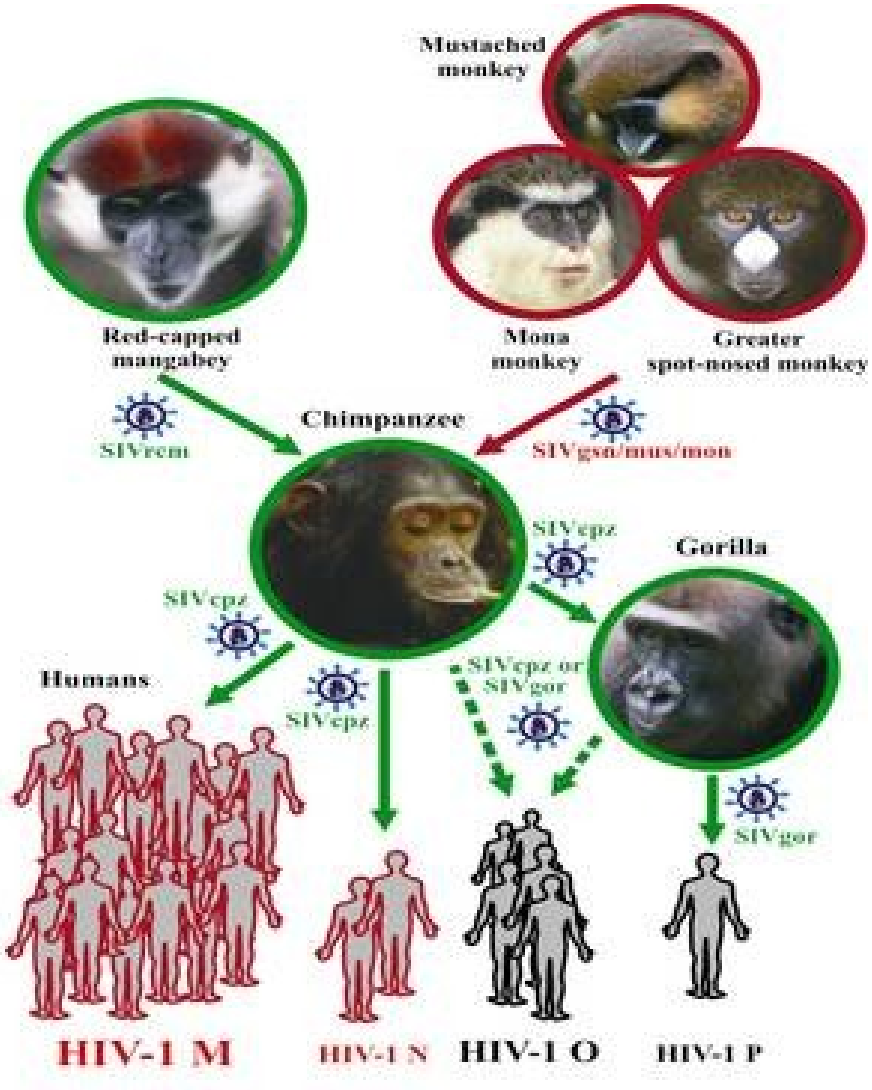
>30 African Old World monkey species are naturally infected with various SIV strains

Absent in Asian Old World monkey species

Symptoms of SIV

- Monkey hosts appear to tolerate heavy viral loads
- No pathogenic effects
- Suggests long coevolution

SIV precursor to HIV

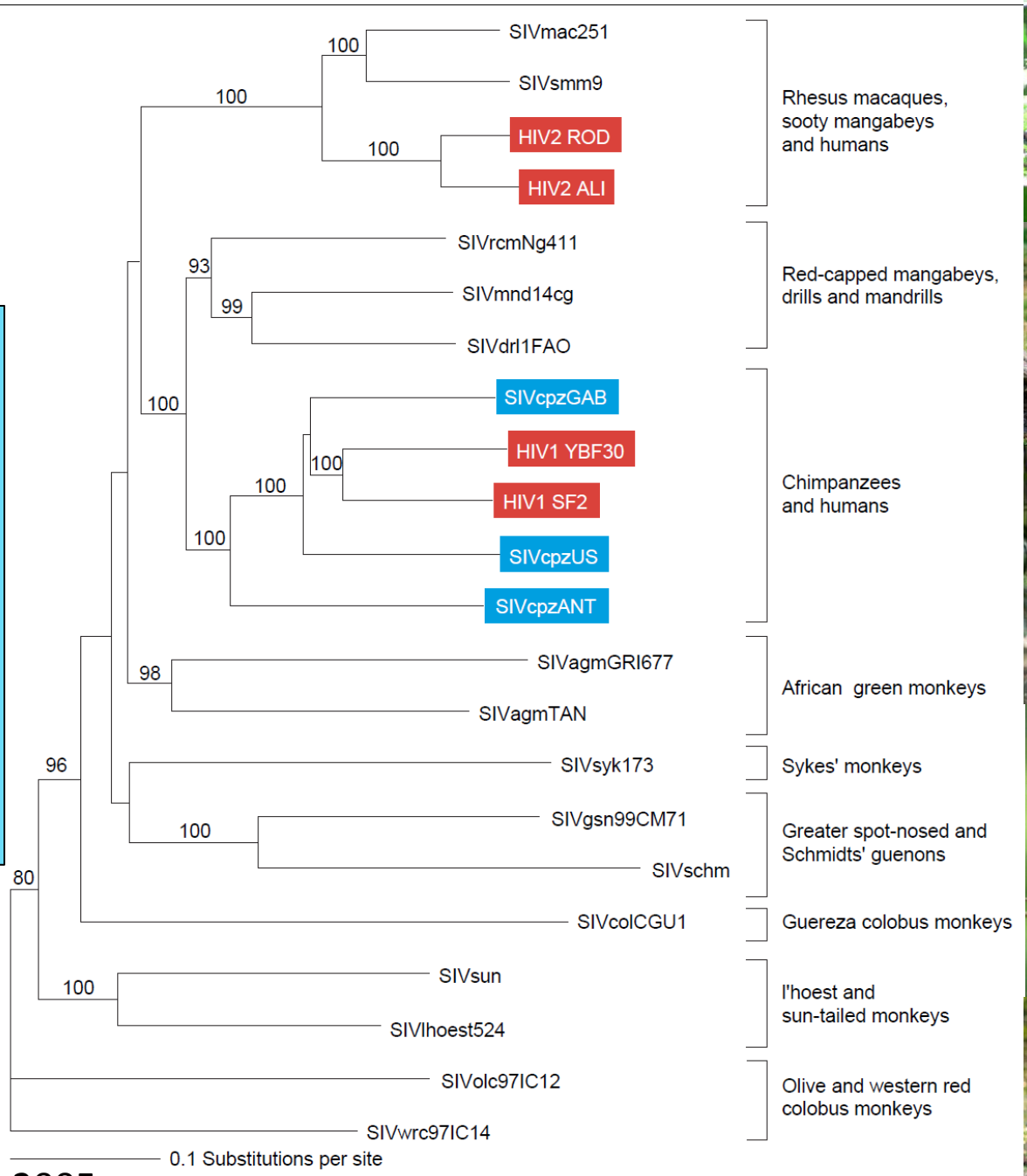


Cross-species transmission

Chimps may have contracted SIV-like infection from Old World Monkeys

Spillover

Zoonotic transfers of SIV to humans have been documented on no fewer than eight occasions



Bontrop and Watkins 2005.

HIV: Where

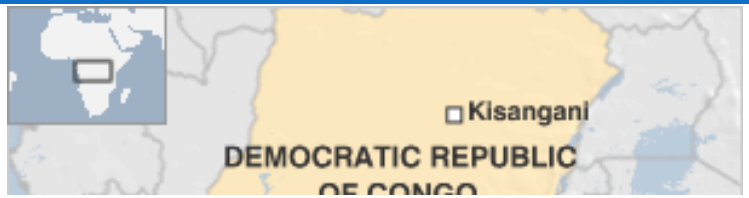

CAMEROON
EQUATORIAL GUINEA
GABON
REPUBLIC OF THE CONGO
DEMOCRATIC REPUBLIC OF THE CONGO
1
2
3
KINSHASA (LEOPOLDVILLE)
1. Humans butcher chimpanzees infected with SIV.
2. The virus is carried by people travelling along the river ...
3. ... to Kinshasa, where the epidemic begins.


Cameroon
International boundary
Province boundary
National capital
Province capital
Railroad
Road
NIGERIA
CHAD
EXTRÊME-NORD
NORD
ADAMAOUA
NORD-OUEST
OUEST
CENTRE
LITTORAL
EST
SUD
CENTRAL AFRICAN REPUBLIC
REPUBLIC OF THE CONGO
EQUATORIAL GUINEA
GABON
SAO TOME AND PRINCIPE
Bight of Biafra
N'Djamena
Yaoundé

HIV: When

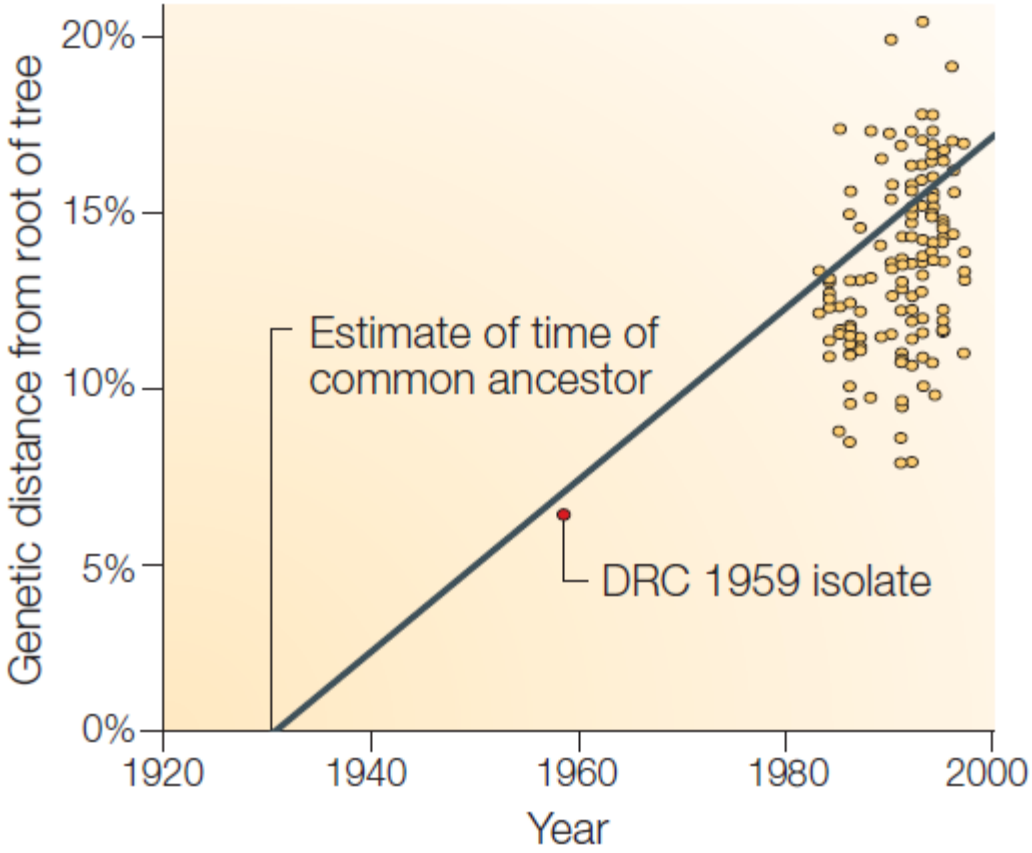
- 2 samples from same year, same city:

1959-60 Kinshasa, DRC.



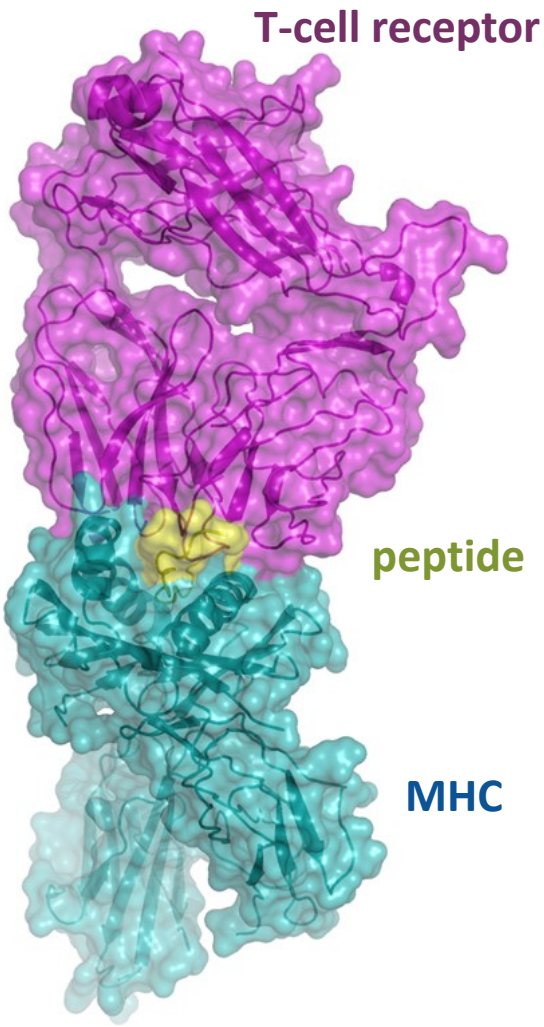
- 12% genetic distance between DRC60 and ZR59 directly demonstrates that there were already at least two distinct clades of HIV in 1960.

- MRCA ~1890-1920



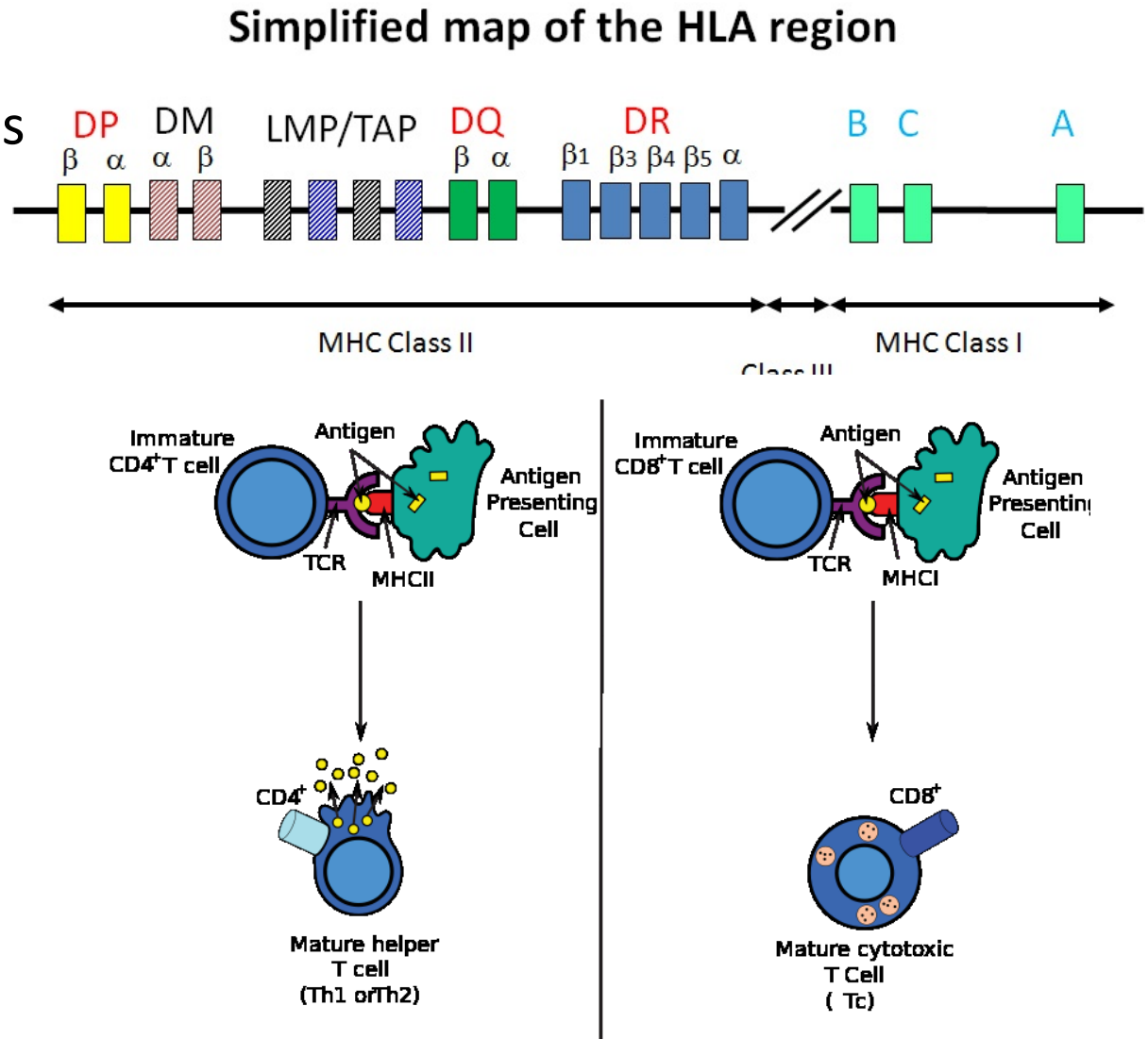
Major Histocompatibility Complex

- MHC Gene Family
  - MHC immune genes of vertebrates
  - Self vs. non-self
  - High diversity



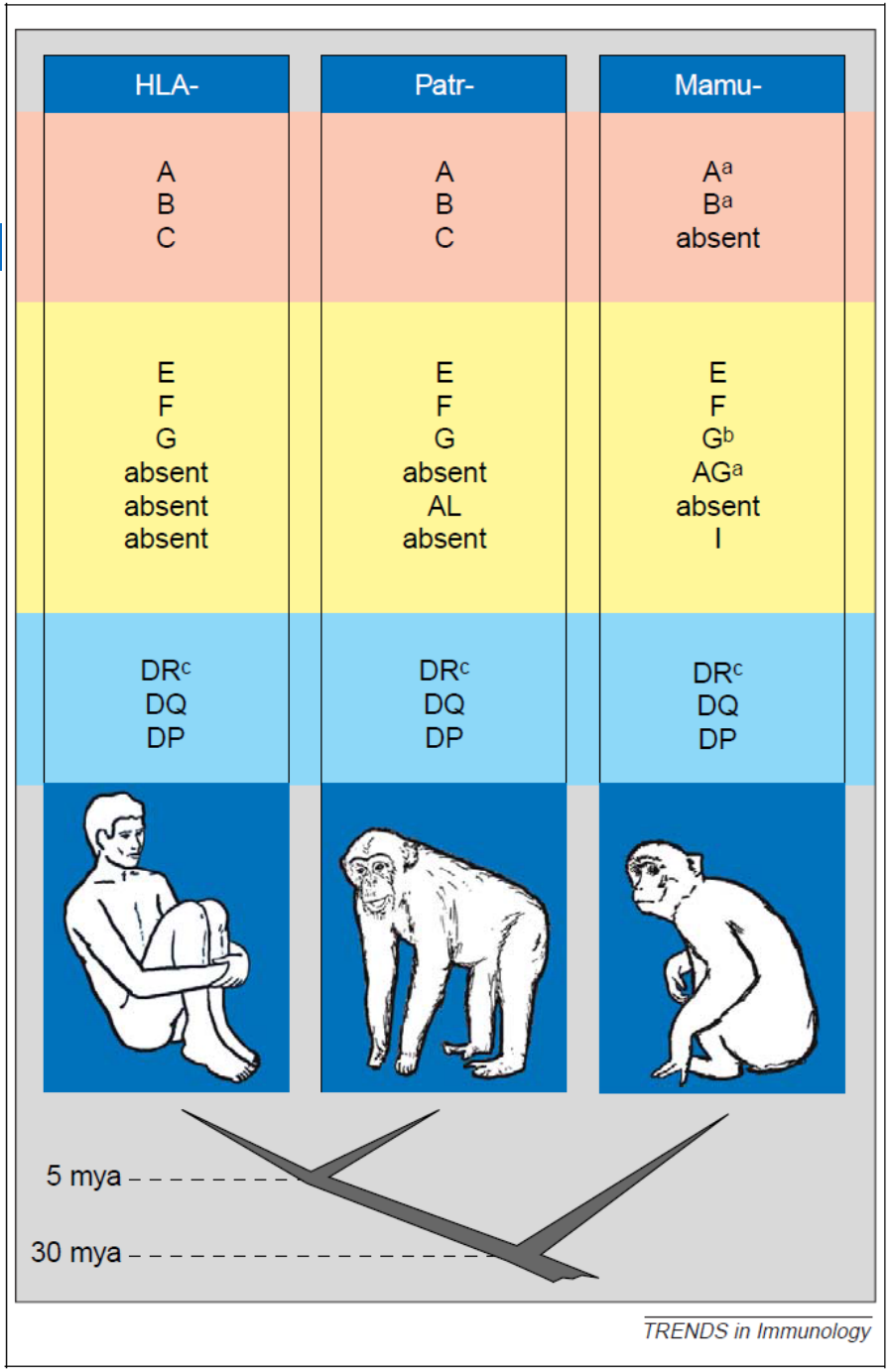
Structure & function of MHC

- Class I
  - Receptors on all cells
  - Intracellular pathogens
  - Cytotoxic “Killer” Tcells
- Class II
  - B-cells and lymphocyes
  - Extracellular pathogens



MHC evolution

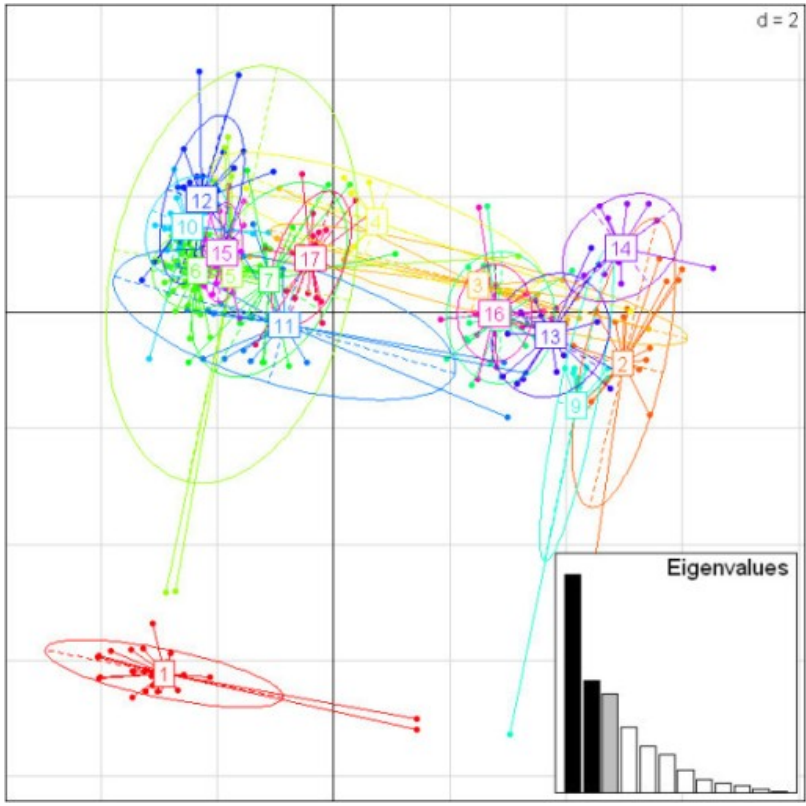
- MHC gene lineages are shared across primates
- Humans and chimps share 98.6% genetic similarity



Bontrop and Watkins 2005.

MHC Supertypes and HIV

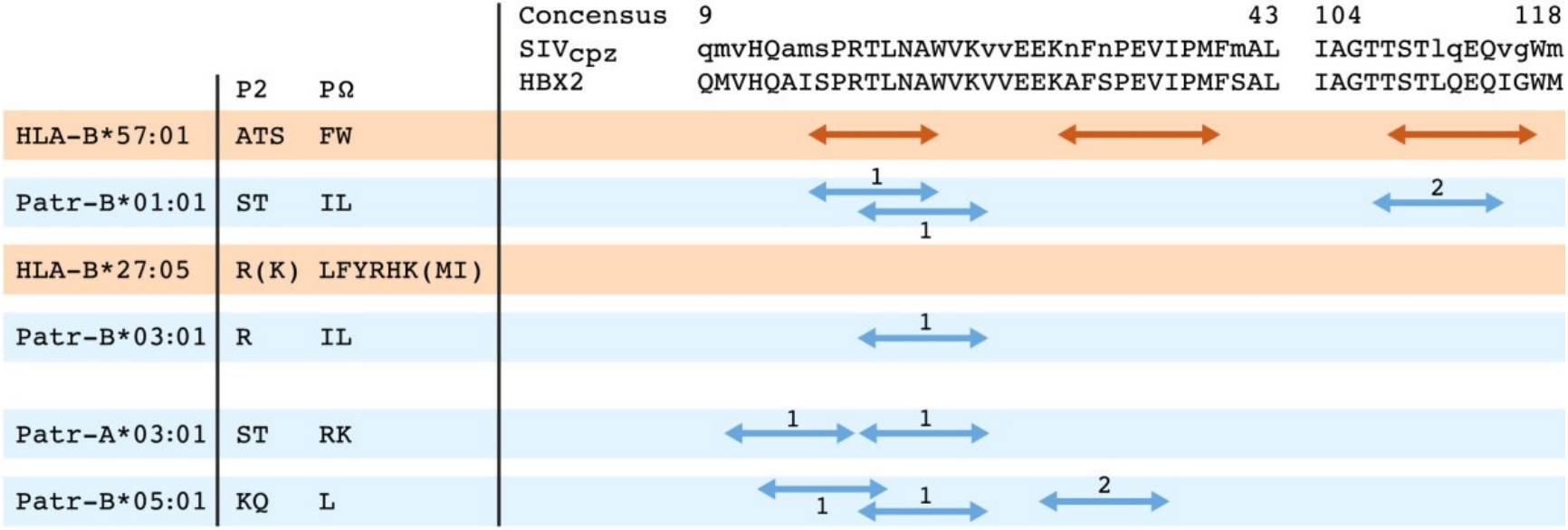
- Binding motifs across alleles that recognize same protein fragments
- Similar supertypes = similar binding affinities
- Short as 1 year or less to a lack of disease progression after more than 35 years and counting in some rare individuals. Supertype associations.



Cross-species protection

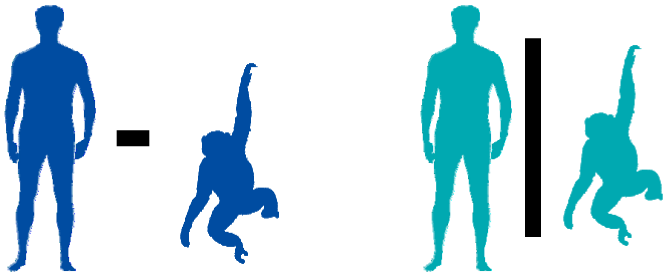
- Some chimpanzee MHC class I-restricted immune responses target conserved epitopes of the HIV-1 virus

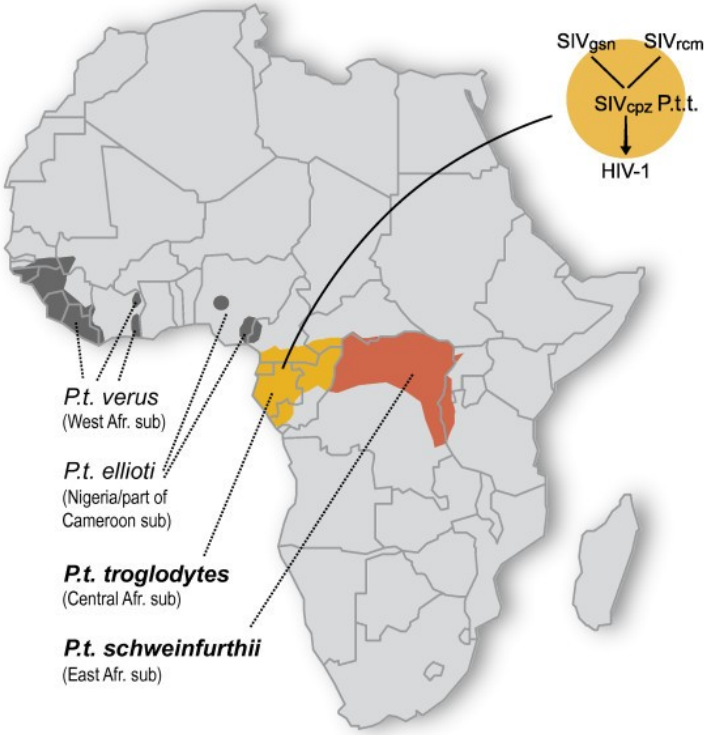
- These *Patr* alleles are characterized by relatively high frequency numbers. Identical viral epitopes are recognized by human long-term nonprogressors

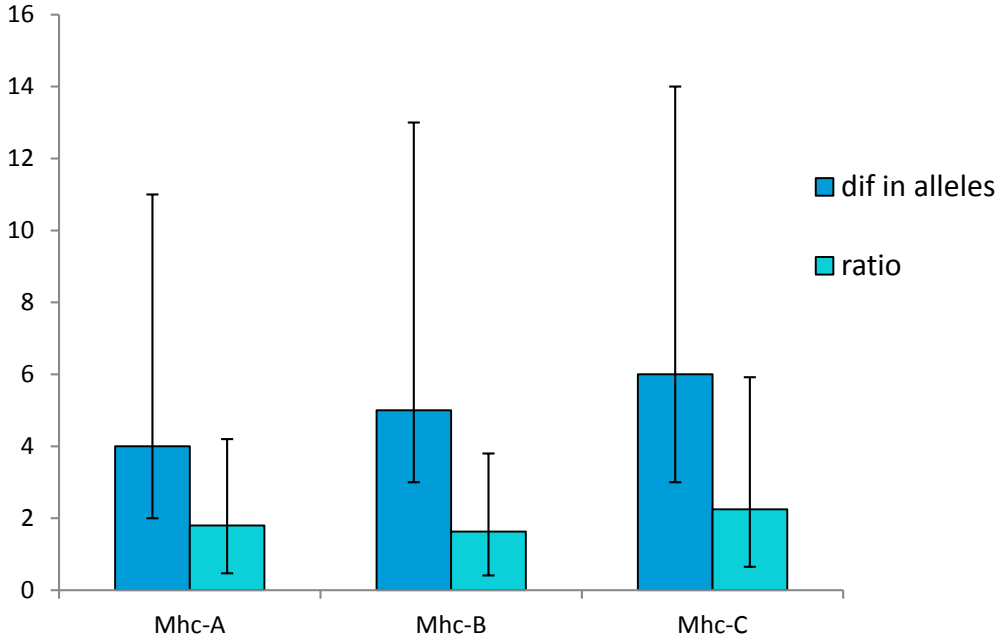


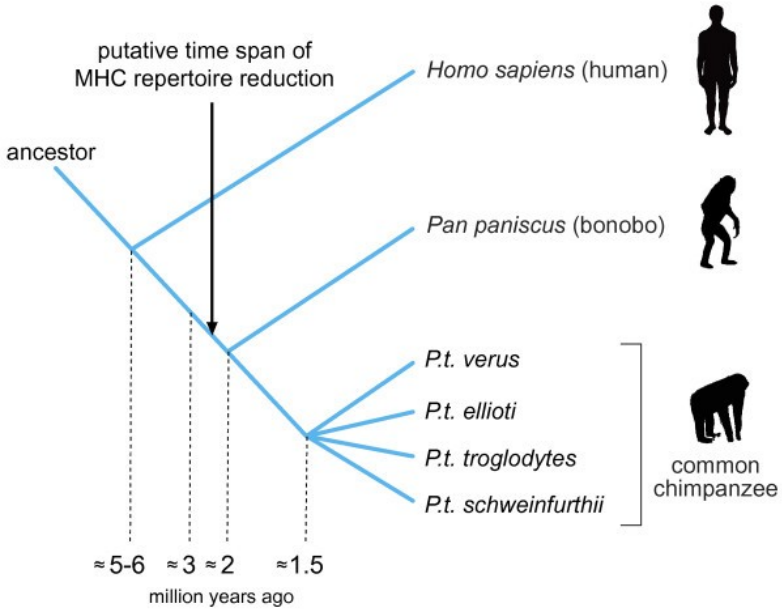
de Groot and Bontrop *Retrovirology* 2013 **10**:53

SIV, HIV and primate MHC resistance



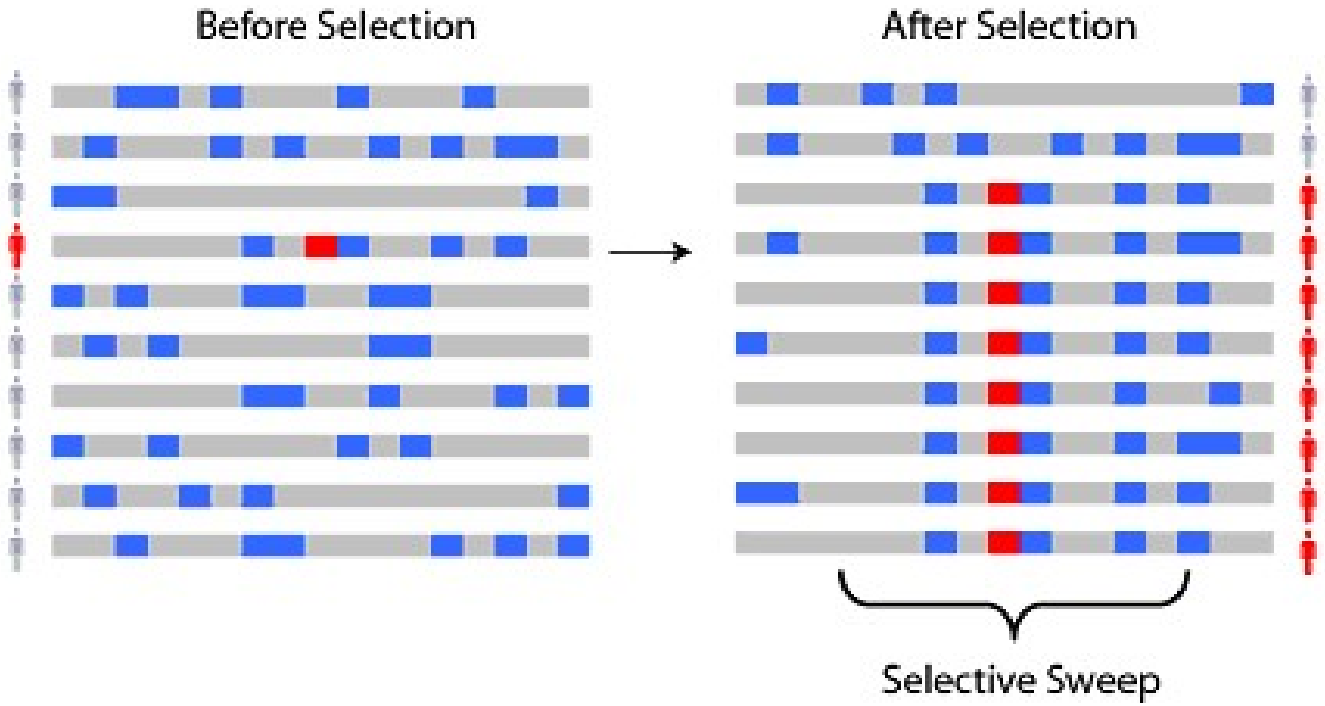




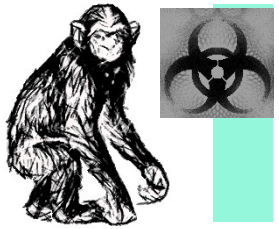


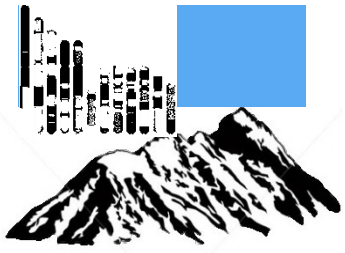
Selective sweeps and genetic hitchhiking

- Evidence of reduced MHC I variation
- Extant variation recognizes/resists HIV-1
- Evidence of lost MHC Class II loci



Outline of talk

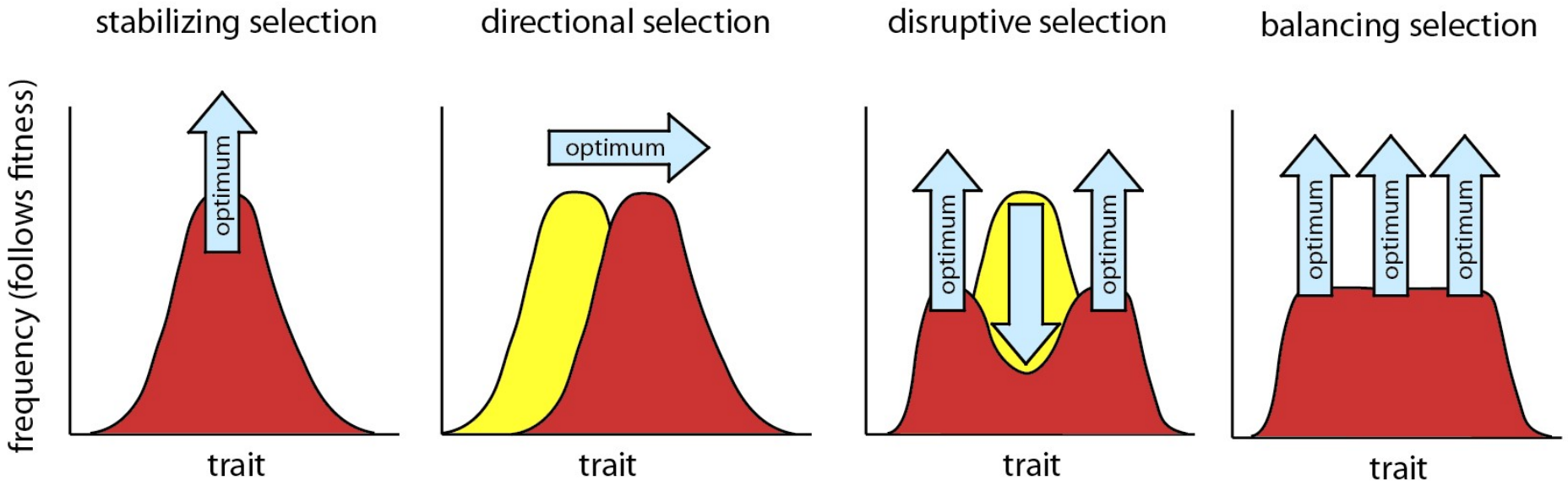




1. **The Chimp and the River**
   - Negative-frequency dependent selection
   - Phylogenetic methods
2. **The Island Fox**
   - Balancing selection
   - Accounting for demography
3. **Men in the Mountains**
   - Positive selection
   - Genome scans

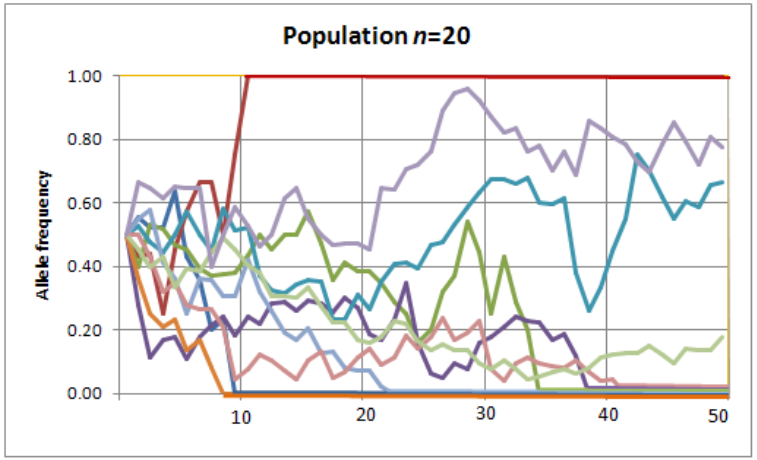
Balancing selection

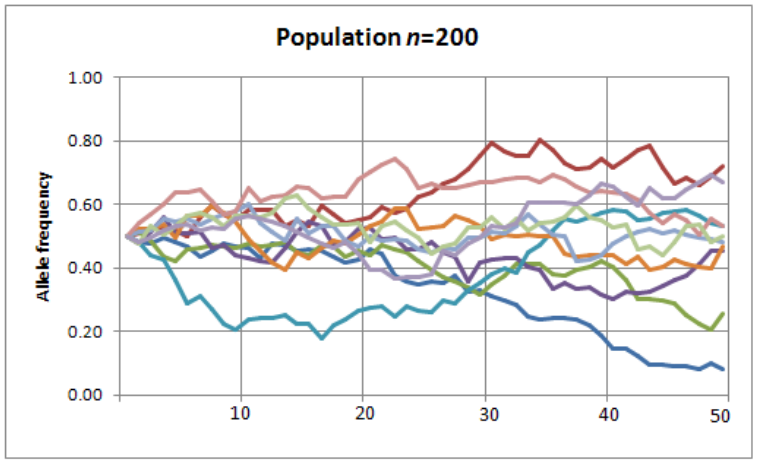
- Selection alters allele frequencies.
- Selection for even “balanced” allele frequencies

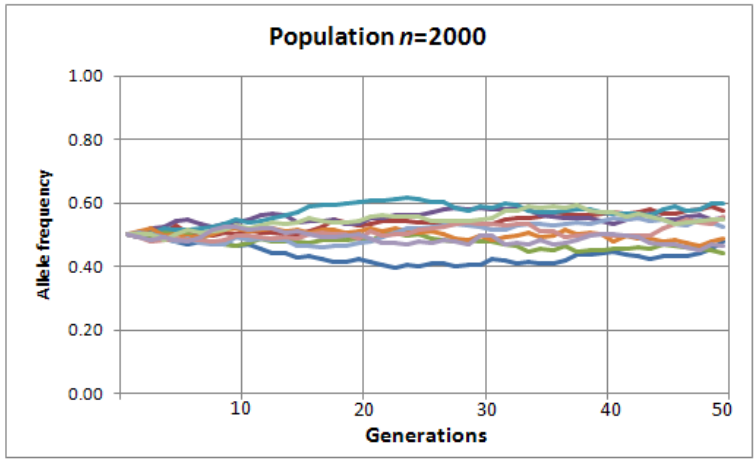


Genetic drift

- Genetic drift alters allele frequencies
- Sampling error with sexually reproducing individuals
- (Effective) population size matters







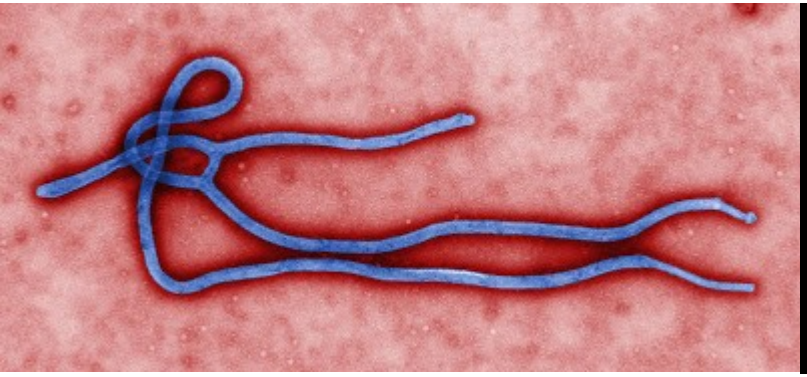
Island Fox

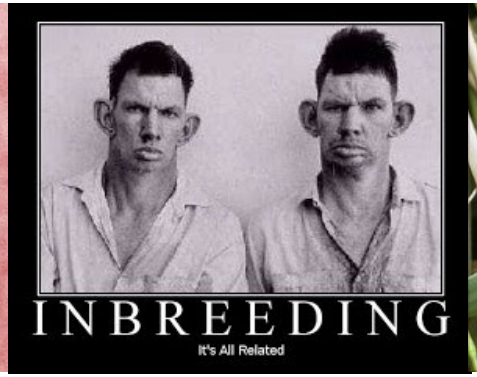
"The San Nicolas Island fox (*Urocyon littoralis dickeyi*) is genetically the most monomorphic sexually reproducing animal population yet reported and has no variation in hypervariable genetic markers."

Aguilar A et al. PNAS 2004;101:3490-3494

Problems with reduced diversity

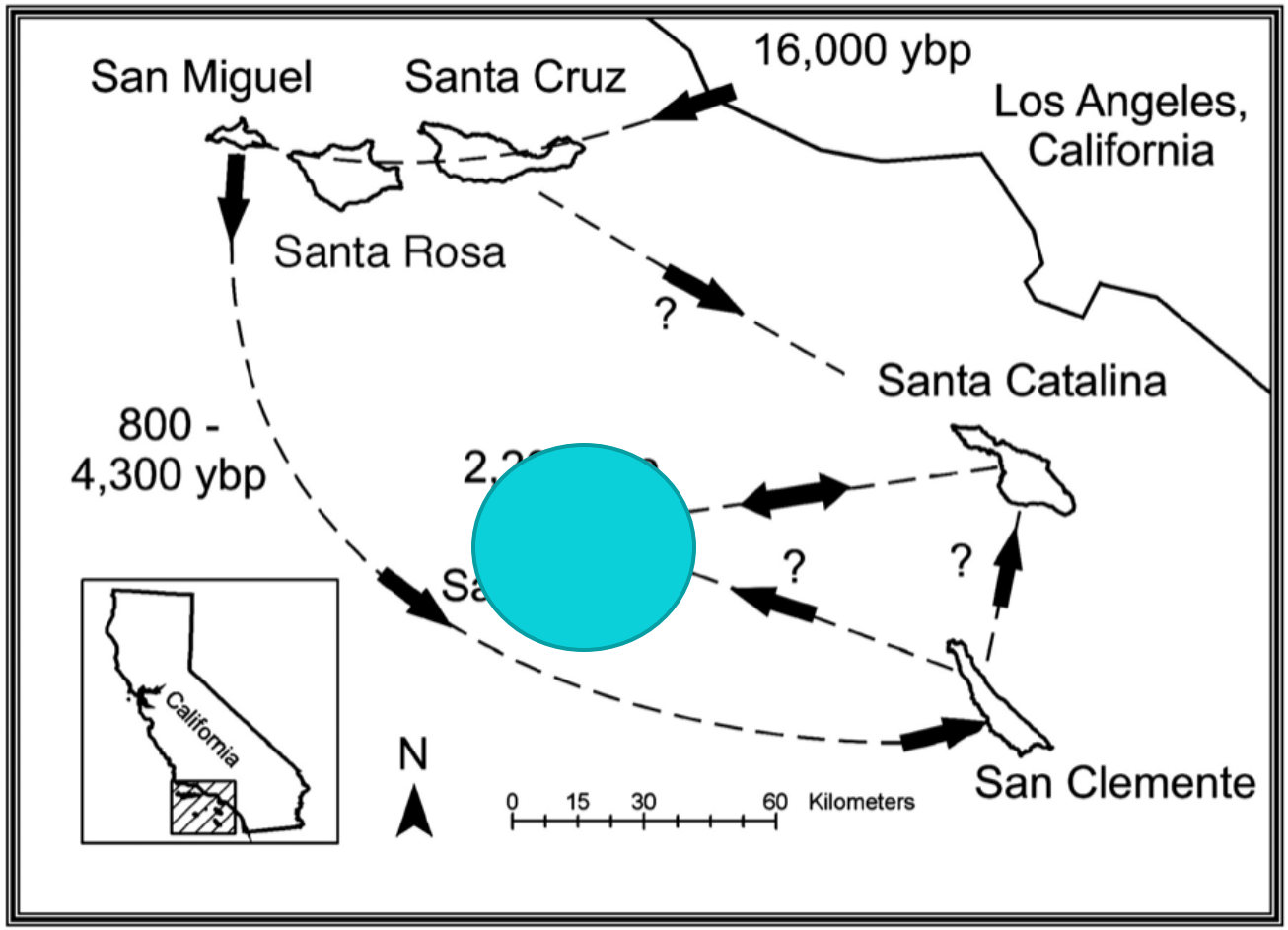
- Lower resistance to pathogens
- Reduced fitness (deleterious recessive alleles unmasked)
- Problems in distinguishing kin from non-kin





Population history

- Levels of genetic variation reflect population size and colonization history
- San Nicolas Island population having the second smallest effective population size and a recent colonization history



Aguilar A et al. PNAS 2004;101:3490-3494

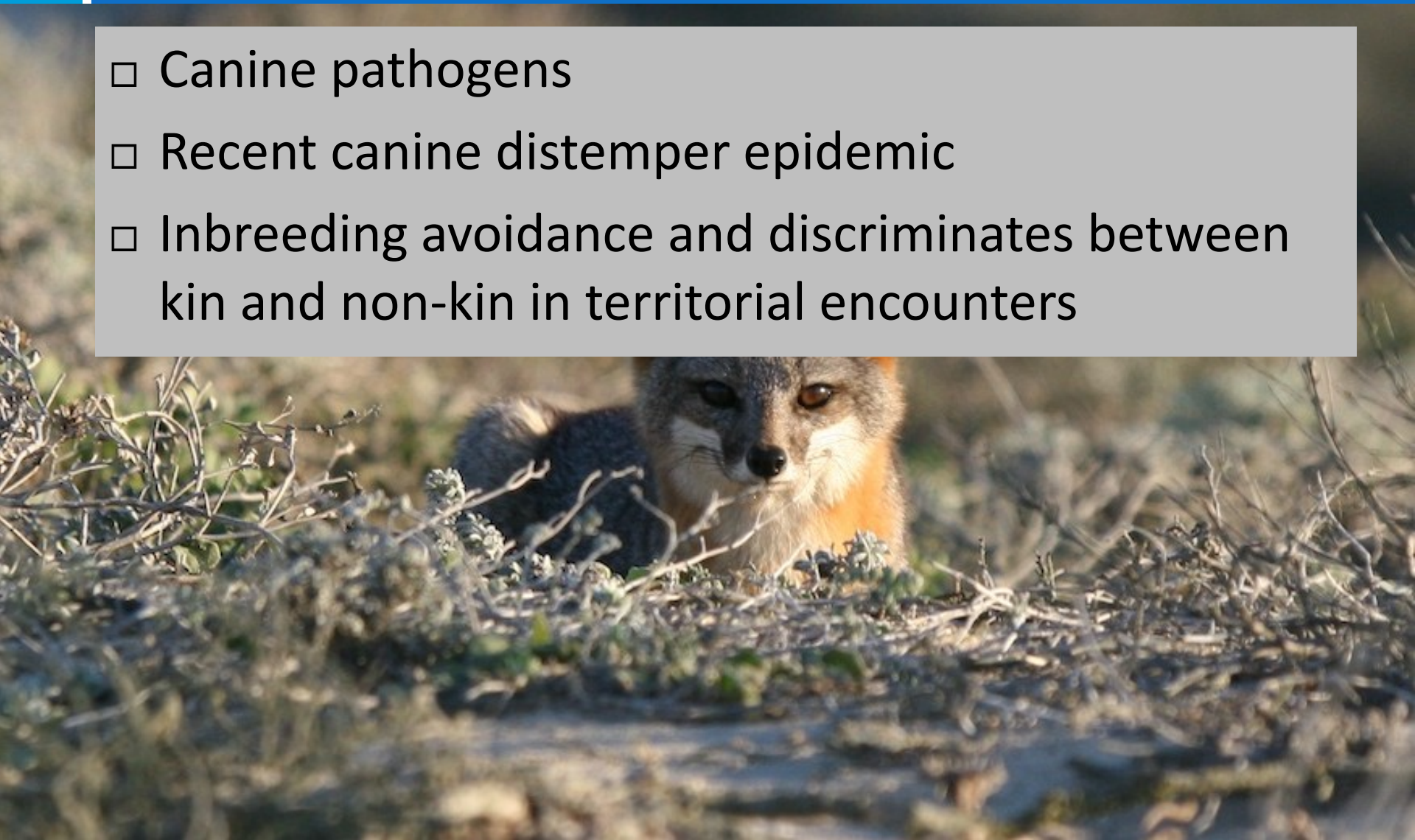
Fox neutral genetic variation

Mean heterozygosity (number alleles)

	N_e	Allozymes	Minisatellites	Microsatellites
San Miguel	163	0.008 (1.1)	0.13	0.11 (1.78)
Santa Rosa	955	0.055 (1.2)	0.34	0.21 (2.56)
Santa Cruz	984	0.041 (1.1)	0.19	0.22 (2.39)
Santa Catalina	979	0.000 (1.0)	0.45	0.36 (2.61)
San Clemente	551	0.013 (1.1)	0.25	0.26 (2.11)

Selective pressures on fox

- Canine pathogens
- Recent canine distemper epidemic
- Inbreeding avoidance and discriminates between kin and non-kin in territorial encounters



Has MHC variation been maintained?

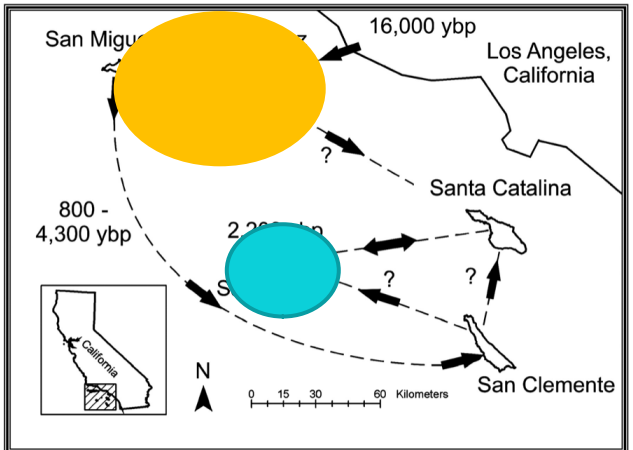
- Objective
  - To determine whether MHC variation has been maintained by natural selection despite the intense genetic drift implied by the genetic monomorphism of neutral genetic markers:
- Quantify MHC variation
  - Assess genetic variability at two class II MHC genes (DRB and DQB) and three class II MHC-linked microsatellite loci.
- Compare MHC variation before and after population separation
  - Compare variation in San Nicolas Island foxes with those on the other Channel Islands
    - estimate levels of MHC variation in populations ancestral to the San Nicolas population
    - account for the influence of population history on levels of MHC variation.
- Simulations
  - Simulations to establish the intensity of selection needed to maintain the observed heterozygosity

Results: MHC variation

Mean heterozygosity (number alleles)

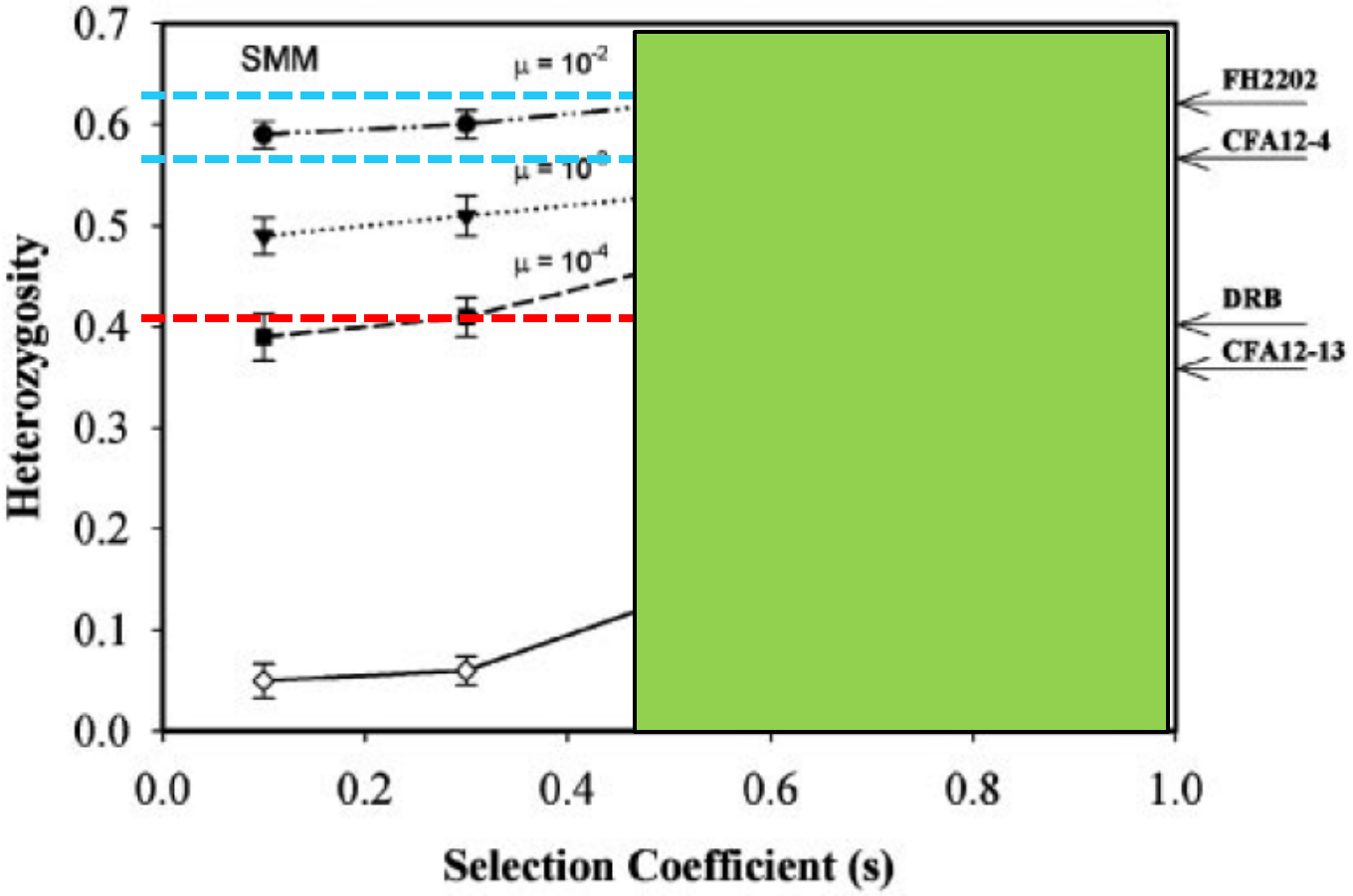
	N_e	n	DRB	DQB	FH2202	CFA12–4	CFA12–13
San Miguel	163	25.8	0.00 (2)	0.00 (1)	0.43 (6)	0.33 (2)	0.50 (4)
Santa Catalina	979	29.0	0.36 (3)	0.55 (4)	0.63 (8)	0.24 (5)	0.37 (3)
San Clemente	551	19.0	0.00 (1)	0.00 (1)	0.68 (5)	0.50 (4)	0.60 (3)

Similar MHC allelic diversity to ancestral populations



Results: Simulations

- SMM: stepwise-mutation model for microsatellites
- IAM: infinite-alleles model for MHC
- μ: mutation rate



Heterozygosity ~ effective population size x mutation rate x selection coefficient

Strength of selection

- LD between DQB and microsats, but not DRB and microsats

- Genetic monomorphism at neutral loci and high MHC variation could arise only through:
  - an extreme population bottleneck of <10 individuals
  - ≈10–20 generations ago
  - unprecedented selection coefficients of >0.5 on MHC loci. (range: 0.05–0.15 in nature)

High periodic selection “rescued” MHC diversity

Critique of story

Heredity (2004) 93, 237–238

www.nature.com/hdy

NEWS AND COMMENTARY

Evolutionary genomics

Foxy MHC selection story

P Hedrick

Heredity (2004) **93,** 237–238. doi:10.1038/sj.hdy.6800539
Published online 28 July 2004

number of organisms, my predisposition is to loudly applaud these findings. However, one needs to be careful in selling an evolutionary story so that it does not become greater that the facts merit.

To provide a perspective for these data, Table 1 gives the observed and

- Lack of LD between DRB and microsats.
- Strong recent selection should show association between microsats near DRB and DRB alleles.

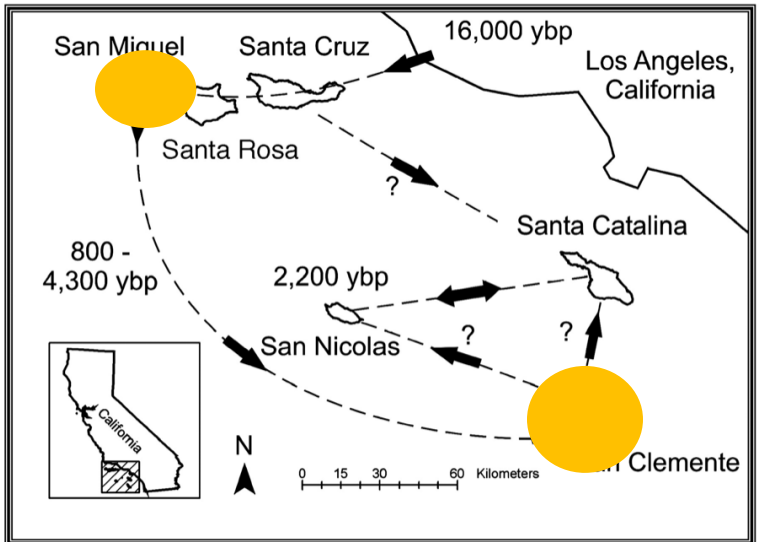
Hedrick 2004. Heredity 93, 237–238

Critique of story

Table 1 The observed (Obs.) and expected (Exp.) heterozygosity for two MHC loci, three microsatellite loci linked to the MHC, and 18 unlinked microsatellite loci in the Island Fox (asterisks indicate benchmarks used in their simulations)

Island	*MHC*				*Microsatellite loci*		
	DRB		*DQB*		*MHC (3)*		*Other (18)*
	Obs.	*Exp.*	*Obs.*	*Exp.*	*Obs.*	*Exp.*	
Santa Rosa	0.16	0.46	0.00	0.00	0.51	0.68	0.21
Santa Cruz	0.14	0.28	0.21	0.40	0.58	0.68	0.22
San Nicolas	0.36*	0.30	0.00	0.00	0.51	0.47	0.00*
Santa Catalina	0.36	0.41	0.55	0.44	0.41	0.59	0.36
Mean	0.17	0.32	0.13	0.14	0.50	0.57	0.19

DRB shows no variation at all on San Miguel or San Clemente Islands



Hedrick 2004. Heredity 93, 237–238

Critique of story

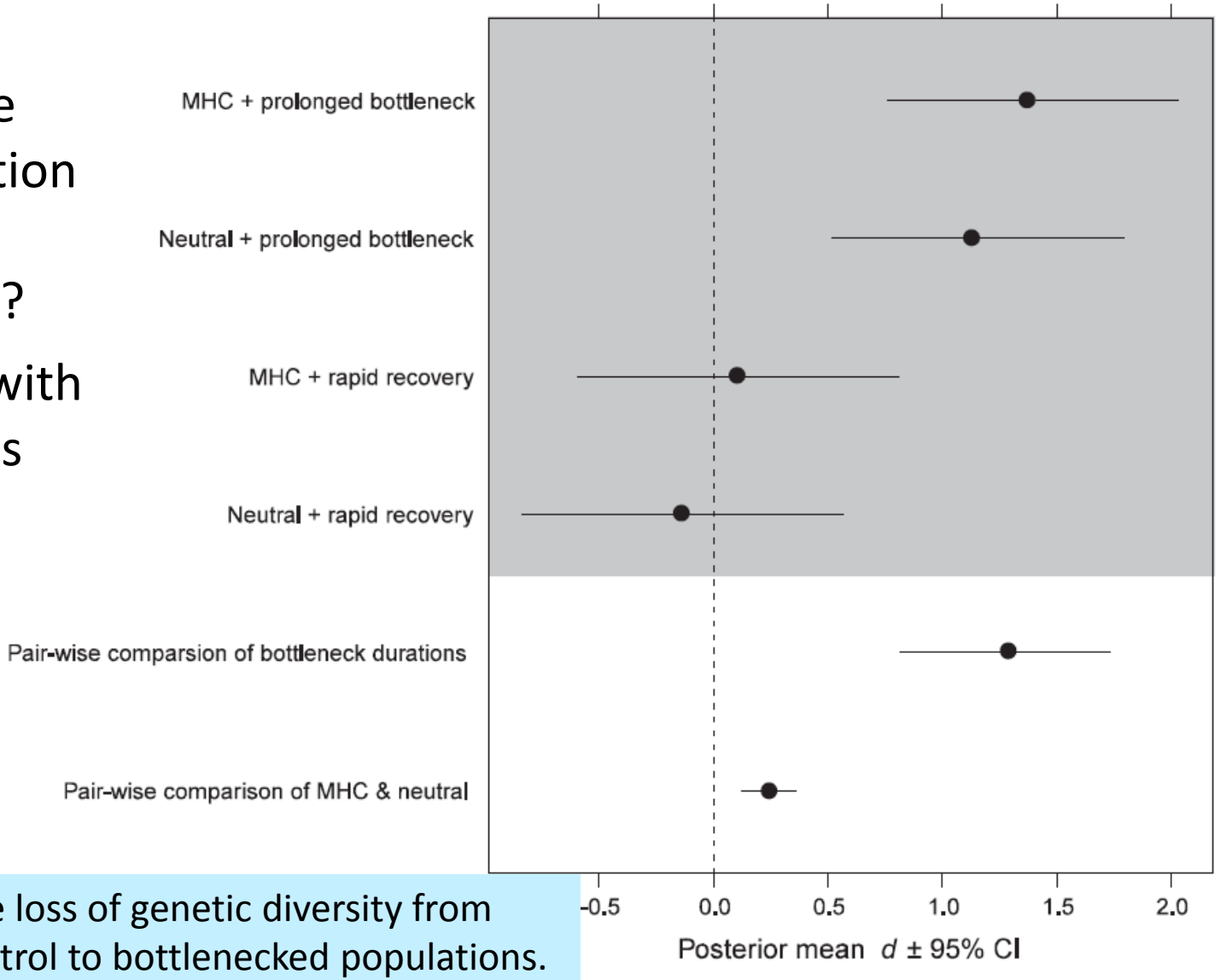
- If DRB were the gene under strong balancing selection, then it is surprising that it shows no variation at all on San Clemente Island, a much larger population.

- If strong selection on DRB, or even other closely linked loci, then the two closely linked MHC microsatellite loci would be expected to still show linkage disequilibrium with DRB.

- Combination of nonselective effects (founder effects) and not-so-extreme balancing selection responsible for empirical results

Meta-analyses and bottlenecks

- Most pops have less MHC variation than neutral variation. Why?
- Meta-analysis with 109 populations (17 studies)

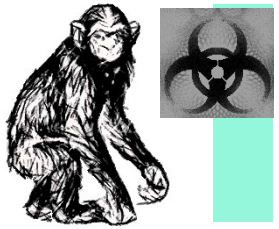


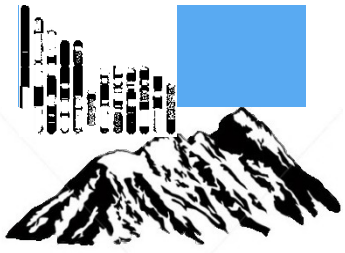
Positive values indicate loss of genetic diversity from pre-bottlenecked / control to bottlenecked populations.

Meta-analyses and bottlenecks

Usually, selection acting on MHC loci prior to a bottleneck event, combined with drift during the bottleneck, will result in overall loss of MHC polymorphism that is ~15% greater than loss of neutral genetic diversity.

Outline of talk

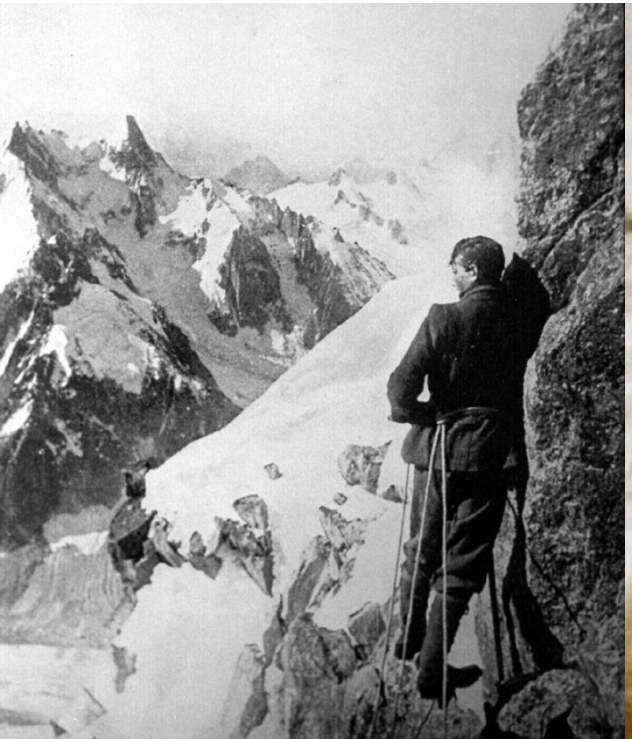




1. **The Chimp and the River**
   - Negative-frequency dependent selection
   - Phylogenetic methods
2. **The Island Fox**
   - Balancing selection
   - Accounting for demography
3. **Men in the Mountains**
   - Positive selection
   - Genome scans

Men of the mountains

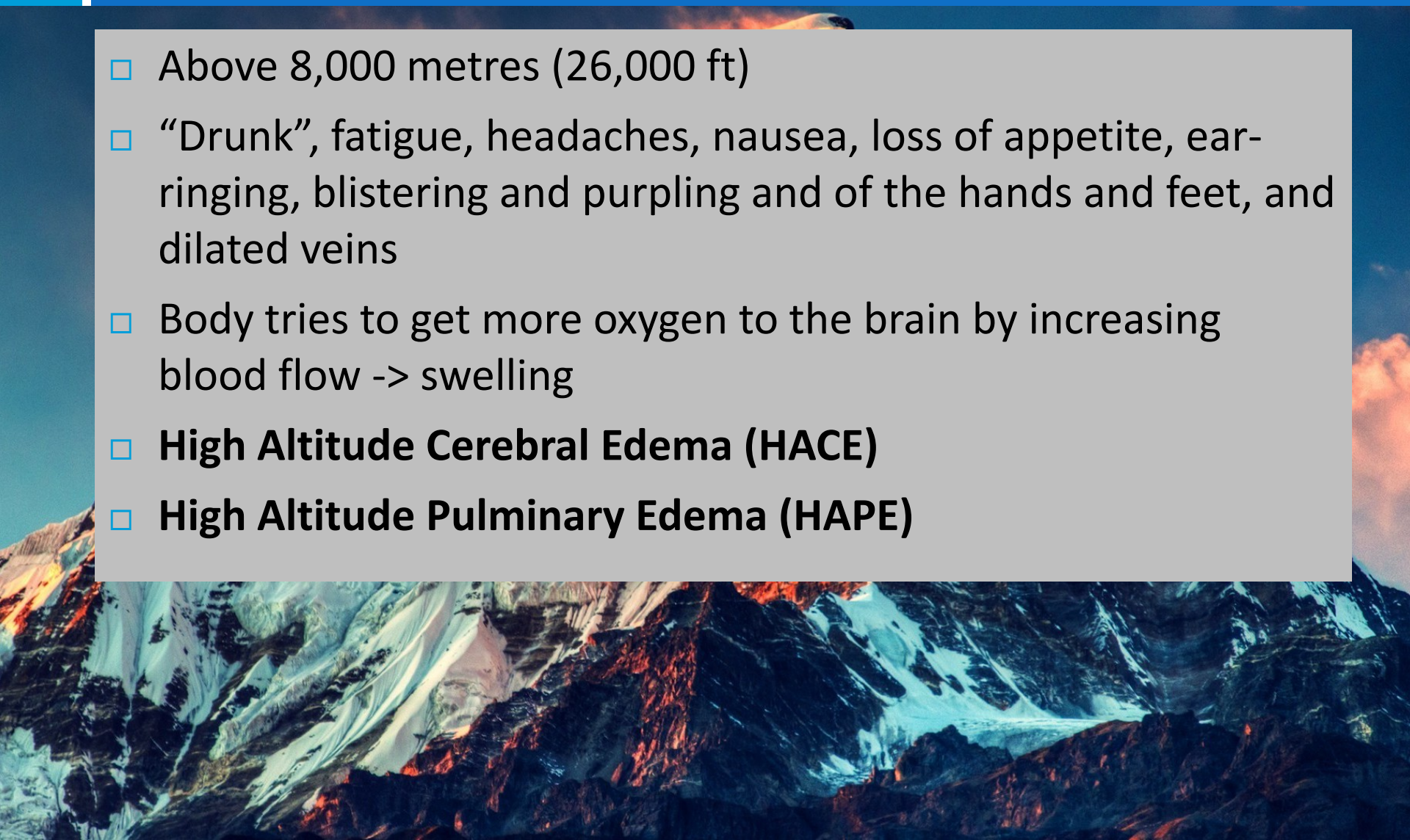
- In 1924 George Mallory and Walter Irvine, 2 first Europeans thought to have achieved summit of Mount Everest, vanished on the descent.



Death on the mountain

- In 1998, Mallory's body was discovered frozen on slope
- Since 1922, over 250 people have died climbing Everest, majority due to events exacerbated by acclimatization issues

The Death Zone



- Above 8,000 metres (26,000 ft)
- “Drunk”, fatigue, headaches, nausea, loss of appetite, ear-ringing, blistering and purpling and of the hands and feet, and dilated veins
- Body tries to get more oxygen to the brain by increasing blood flow -> swelling
- **High Altitude Cerebral Edema (HACE)**
- **High Altitude Pulminary Edema (HAPE)**

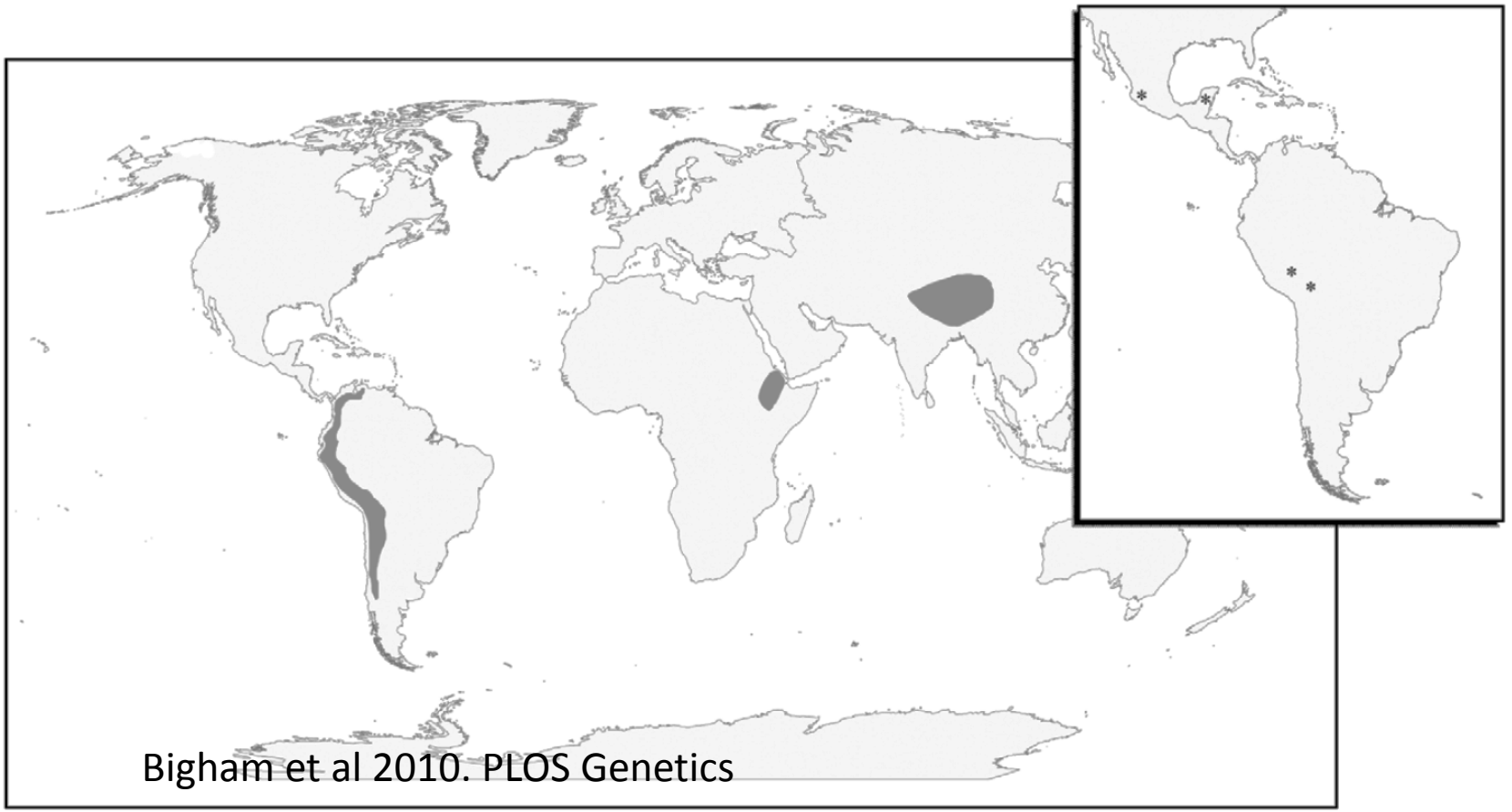
High altitude adaptations

- Decreased oxygen availability (>2,500 m)
- Decreased barometric pressure

- Physiological changes
  - increased lung volumes,
  - increased breathing
  - higher resting metabolism
  - hemoglobin changes

Geography of human adaptation to high altitude

- Andean Altiplano, Ethiopian Highlands, Tibetian Plateau
- Populated 11,000 - 25,000 years ago



Bigham et al 2010. PLOS Genetics

Genome scans for selection

- Goal: Identify candidate genes for high-altitude adaptation based on signatures of positive selection in Tibetian and Andean populations

- What are we looking for?
- How do we know if the region is under selection vs random variation between individuals?

Design of study

1. Contrast high-altitude populations with low-altitude population controls
    1. Andean vs Mesoamerican and East Asian
    2. Tibetan vs European and East Asian
2. Use 4 different complimentary tests of natural selection
3. Compare independent high-altitude population results

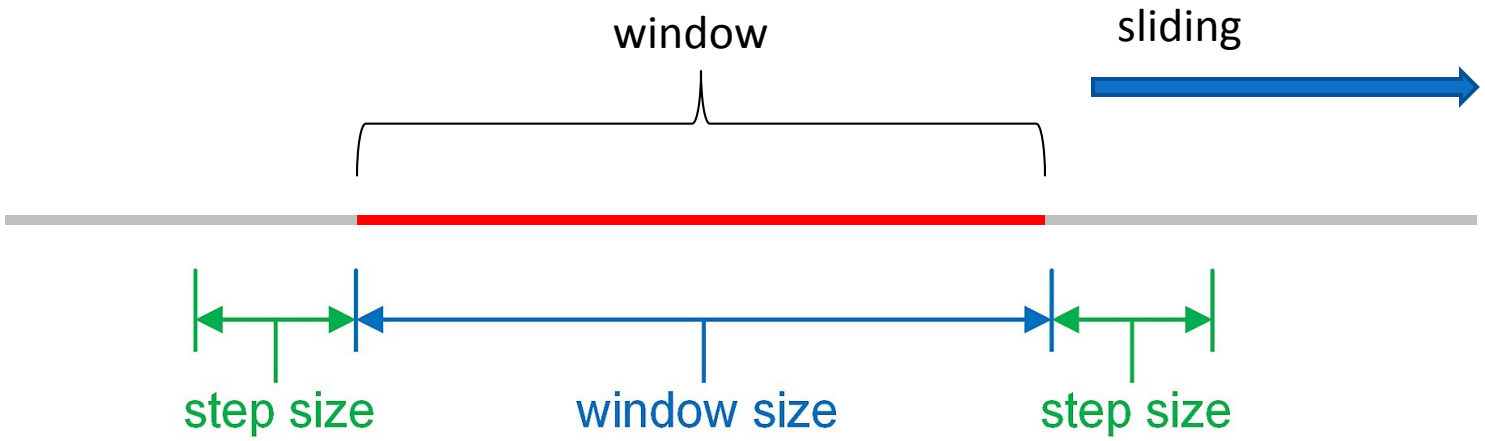
Tests of natural selection

- 1) natural-log ratio of heterozygosity (lnRH)
- 2) standardized difference of Tajima's D
- 3) whole genome long range haplotype (WGLRH)

- Statistical significance determined using genome-wide empirical distributions generated by data.

1) Ratio of heterozygosity (lnRH)

- Natural log of ratio of heterozygosity between 2 pops of interest (High vs Low altitude pops)
- Sliding window of 100,000bp in 25,000bp increments along a chromosome



Negative lnRH values = regions with reduction in variation in high altitude population

Tajima's D

1	2	3	4	5	6	7	8
0	1	0	0	0	0	0	0
0	0	0	1	0	1	1	0
0	0	0	1	0	1	1	0

- Under neutrality:

$$E[\pi] = \theta = E\left[\frac{S}{\sum_{i=1}^{n-1}\frac{1}{i}}\right] = 4N\mu$$

$$\mathbf{D} = \frac{(\mathbf{E(\pi)\text{-}E(S)})}{\mathbf{stdev(E(\pi)\text{-}E(S))}}$$

- (Average #pairwise polymorphisms-standardized #segregating sites)/stdDev(d)
- Average Heterozygosity = # of Segregating sites
- E(π)= (4+0+4)/3 = 2.67
- E(S) = 4 sites/(1/1+1/2) = 2.67
- D = 2.67-2.67/sqrt[Var(d)] = 0, Neutrality
- If AvgHet > Segregating sites, D>0: Intermediate freq alleles, **Balancing selection** or recent pop bottleneck that removed rare alleles
- If AvgHet < Segregating sites, D<0: High freq of singletons, **Positive or purifying selection, selective sweep**

Worked D examples

$$D = \frac{(E(\pi)-E(S))}{\text{stdev}(E(\pi)-E(S))}$$

- Number of pairs = n(n-1)/2
- = 4(3)/2 = 12/2 = 6

- **Blue Table**
- π=(5+3+2+2+3+3)= 18/6 = 3
- S = 5 sites/(1/1+1/2+1/3) = 5/(1.83) = 2.73
- D = 3-2.73 = 0.27 **D>0**

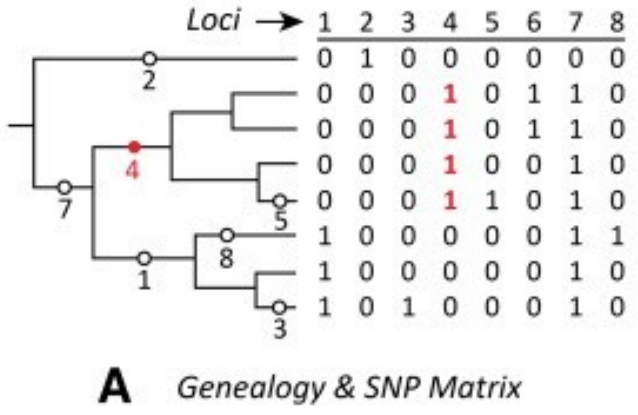
	1	2	3	4	5	6	7	8
A	0	1	0	0	1	0	0	0
B	0	0	0	1	0	0	1	1
C	0	0	0	0	0	0	1	0
D	0	1	0	1	0	0	0	0

- **Green Table**
- π=(5+5+5+0+0+0) = 15/6 = 2.50
- S = 5 sites/(1/1+1/2+1/3) = 2.73
- D = 2.5-2.73 = -0.23 = **D<0**

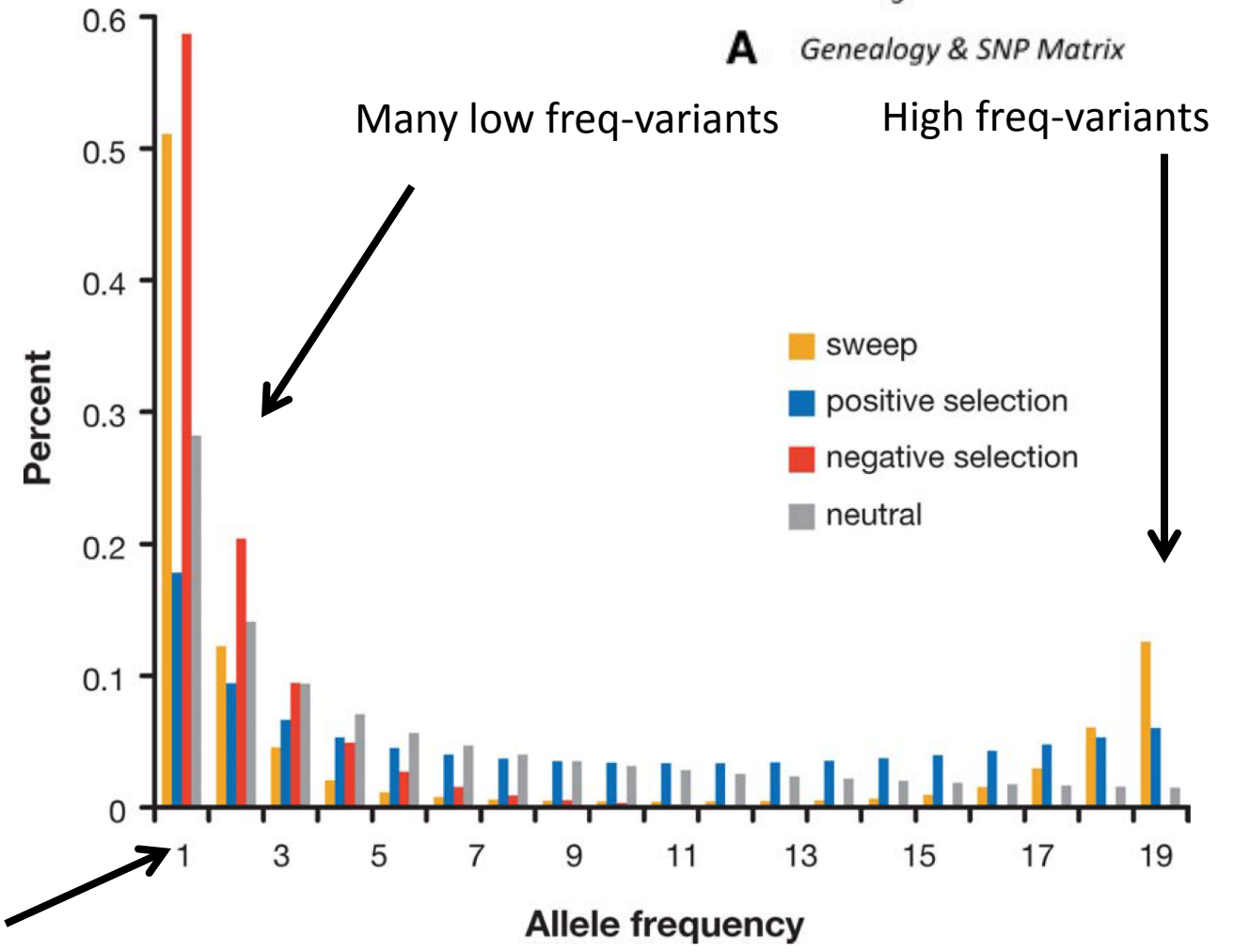
	1	2	3	4	5	6	7	8
A	1	1	1	1	1	0	0	0
B	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0

Must know the standard deviation to determine significance

Frequency spectrum



- In a standard neutral model
  - Random mating
  - Constant populatior size
  - No population subdivision



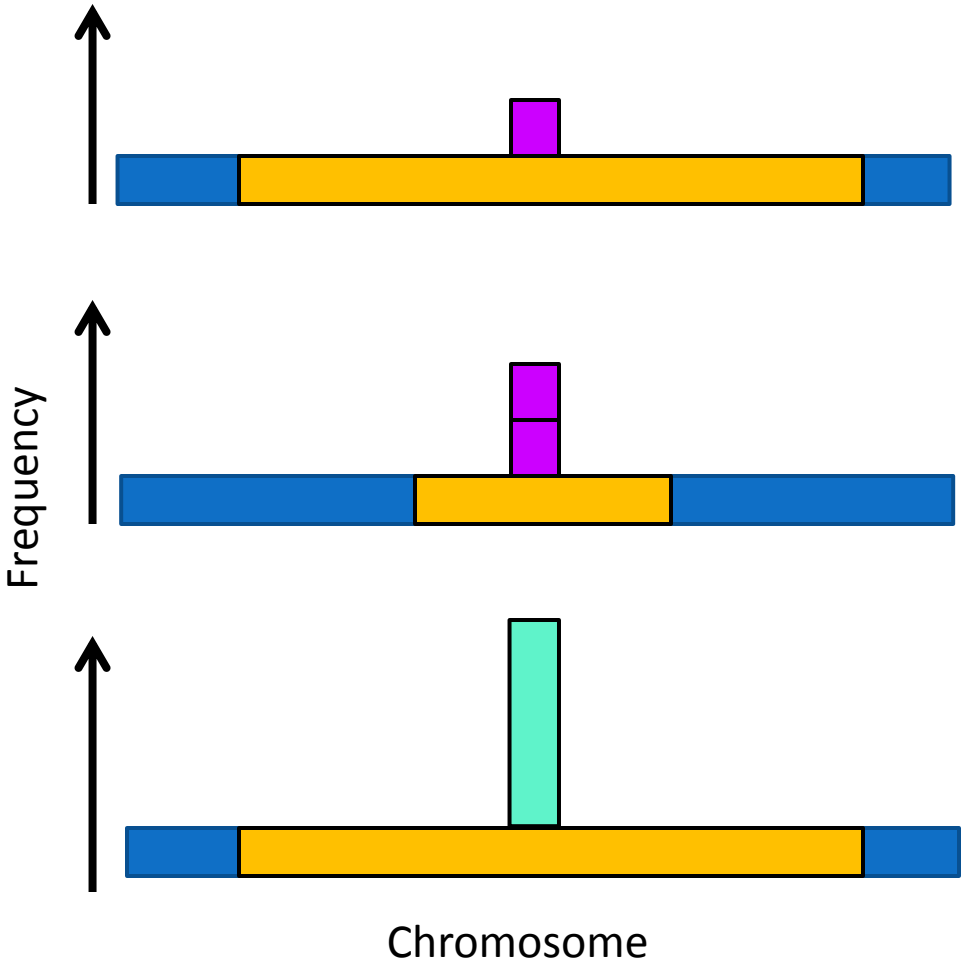
2) Standardized difference in D

Standardized difference of D = $\frac{(D\,i_{High} - Di_{Low}) - \mu\,(D_{High} - D_{Low})}{SD(D_{High} - D_{Low})}$

- Di = Tajima's D in sliding window
- μ = mean Tajima's D for all windows
- High = Andean or Tibetian population
- Low = Control low altitude population

Negative standardized D = regions under selection in high altitude population controlling for demographic events

3) Whole genome long range haplotype (WGLRH)

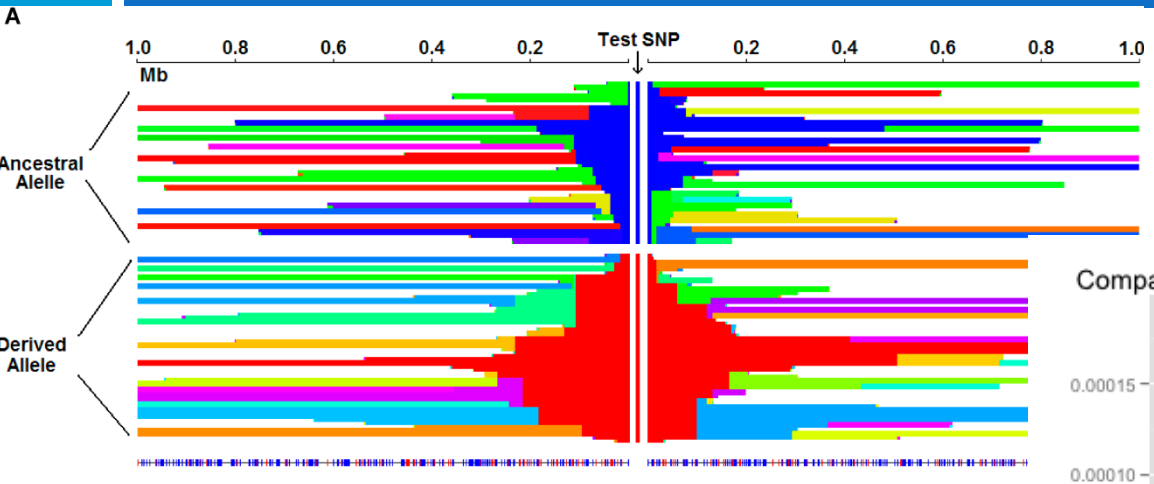


Young allele (neutral)
- Low frequency
- Long range LD
- No time for recombination

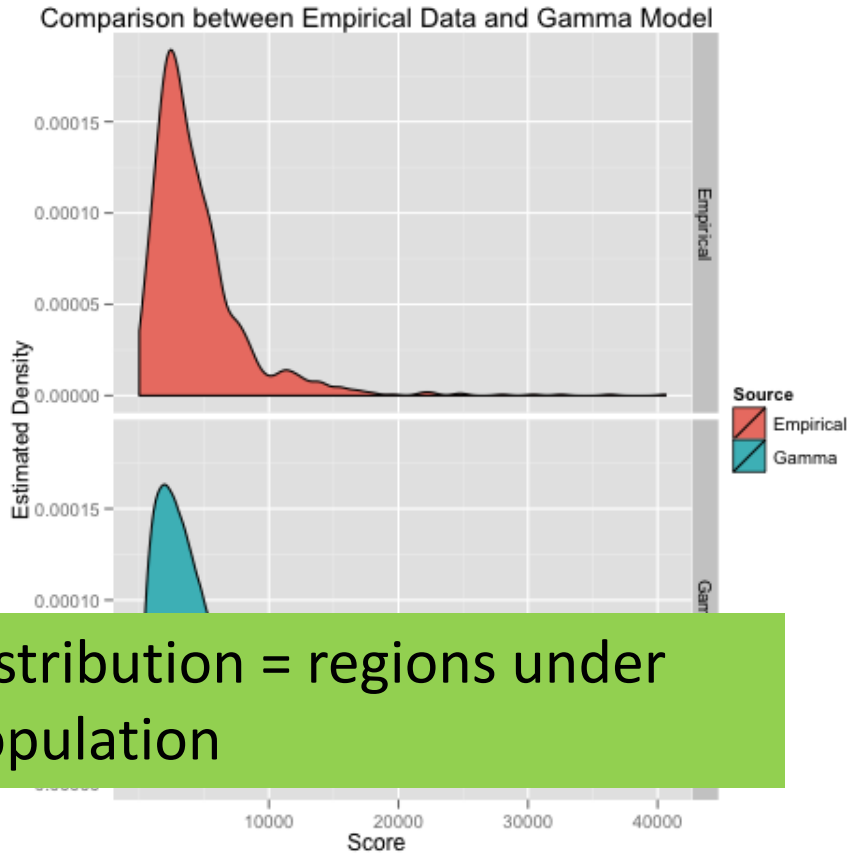
Old allele (neutral)
- Low or high frequency (drift)
- Short range LD
- Lots of recombination

Young **selected** allele
- High frequency
- Long-range LD
- Hitch-hiking of linked sites

Long range haplotype

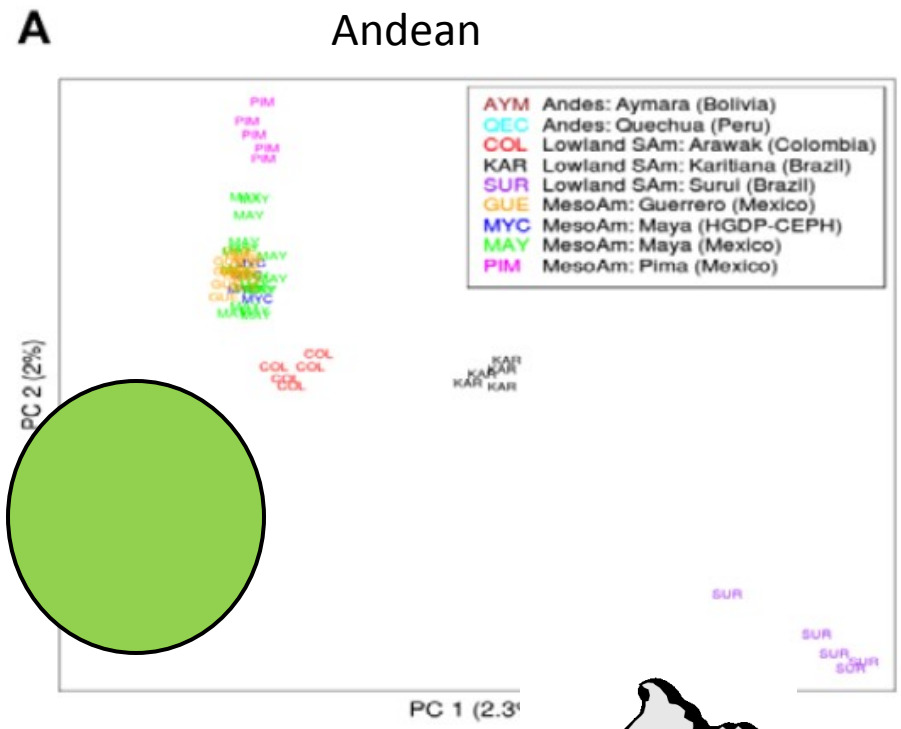


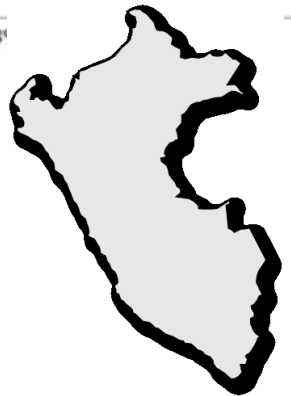
Compare Relative Extended Haplotype Homozygosty to flexible gamma distribution parameterized with maximum likelihood methods from rest of dataset

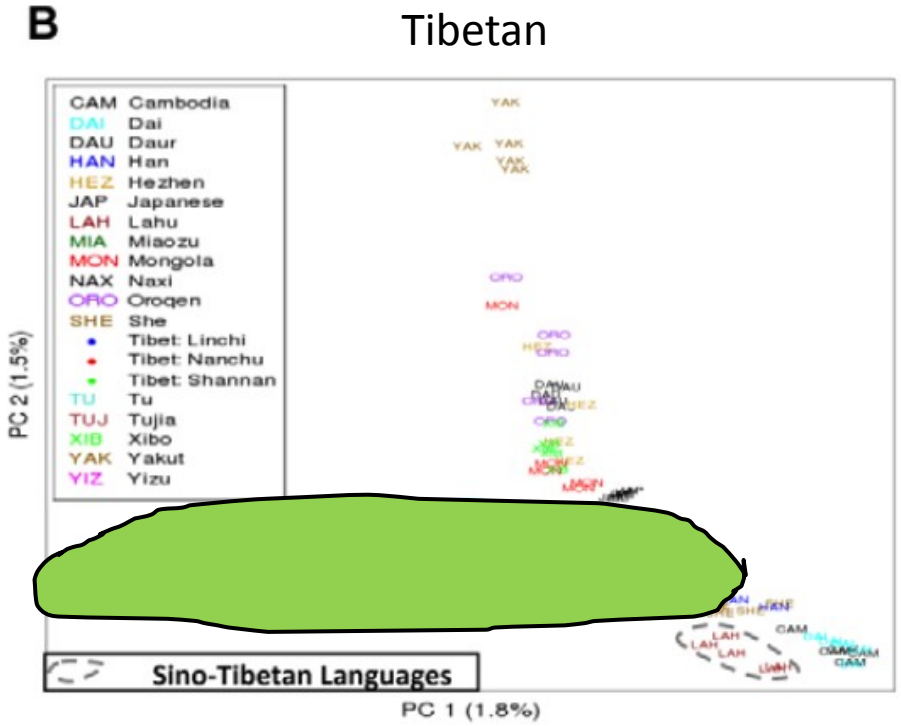


Values in upper 5% tail of gamma distribution = regions under positive selection in high altitude population

Results: individual ancestry estimates

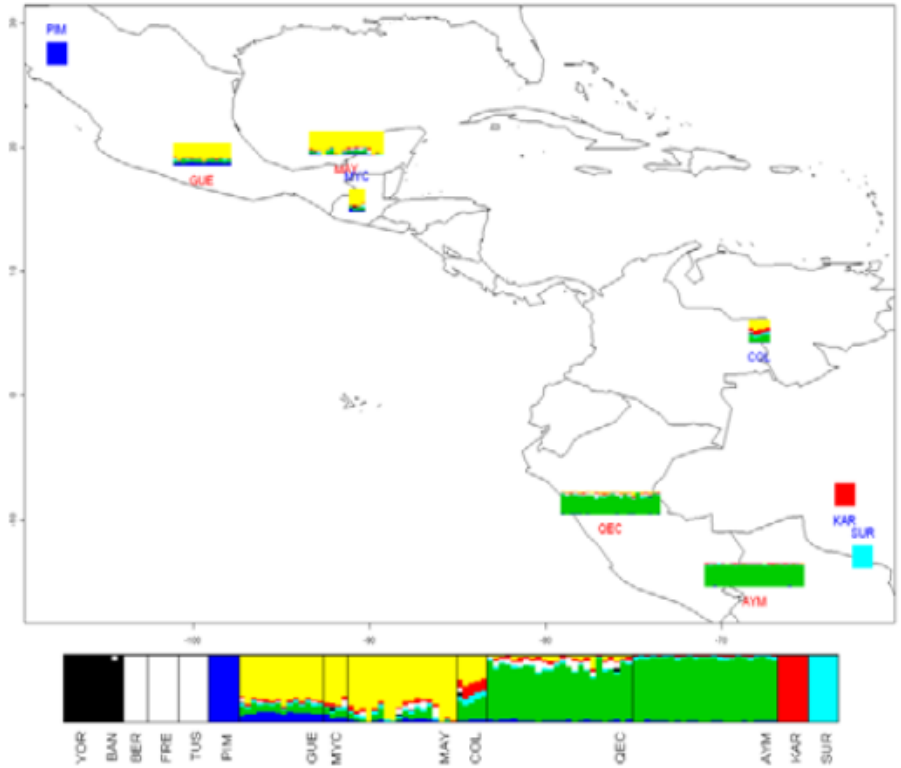




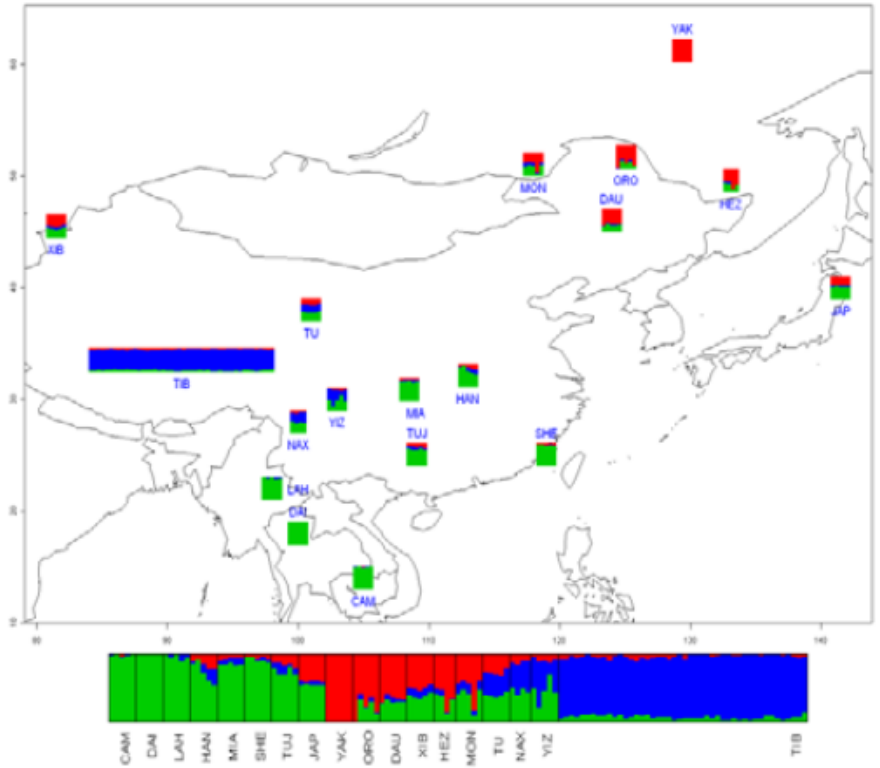


Results: population stratification

Andean



Tibetan



Results: Genome scans

Table 1. Significant SNPs or SNP windows in Andeans and Tibetans for $P_E \leq 0.05$ and $P_E \leq 0.01$.

Population	Test	Autosomes	$P_E = 0.05$	$P_E = 0.01$	X	$P_E = 0.05$	$P_E = 0.01$
Andean	LSBL	856,231	42,812	8,562	36,160	1,808	362
	ln*RH*	106,163	5,308	1,062	5,869	293	59
	D	106,109	5,305	1,061	5,862	293	59
	WGLRH	69,226	178	NA	271	0	NA
Tibetan	LSBL	845,054	42,253	8,451	36,031	1,802	360
	ln*RH*	106,140	5,307	1,061	5,869	293	59
	D	106,093	5,305	1,061	5,862	293	59
	WGLRH	79,938	436	NA	1046	2	NA

Autosomes and the X chromosome are listed separately.
doi:10.1371/journal.pgen.1001116.t001

- MANY significant SNPs for both populations, varying by test
- Strength of selection, time since selection, and recombination background all affect signal and test sensitivity

Results: Genetic variation at cellular oxygen sensing gene

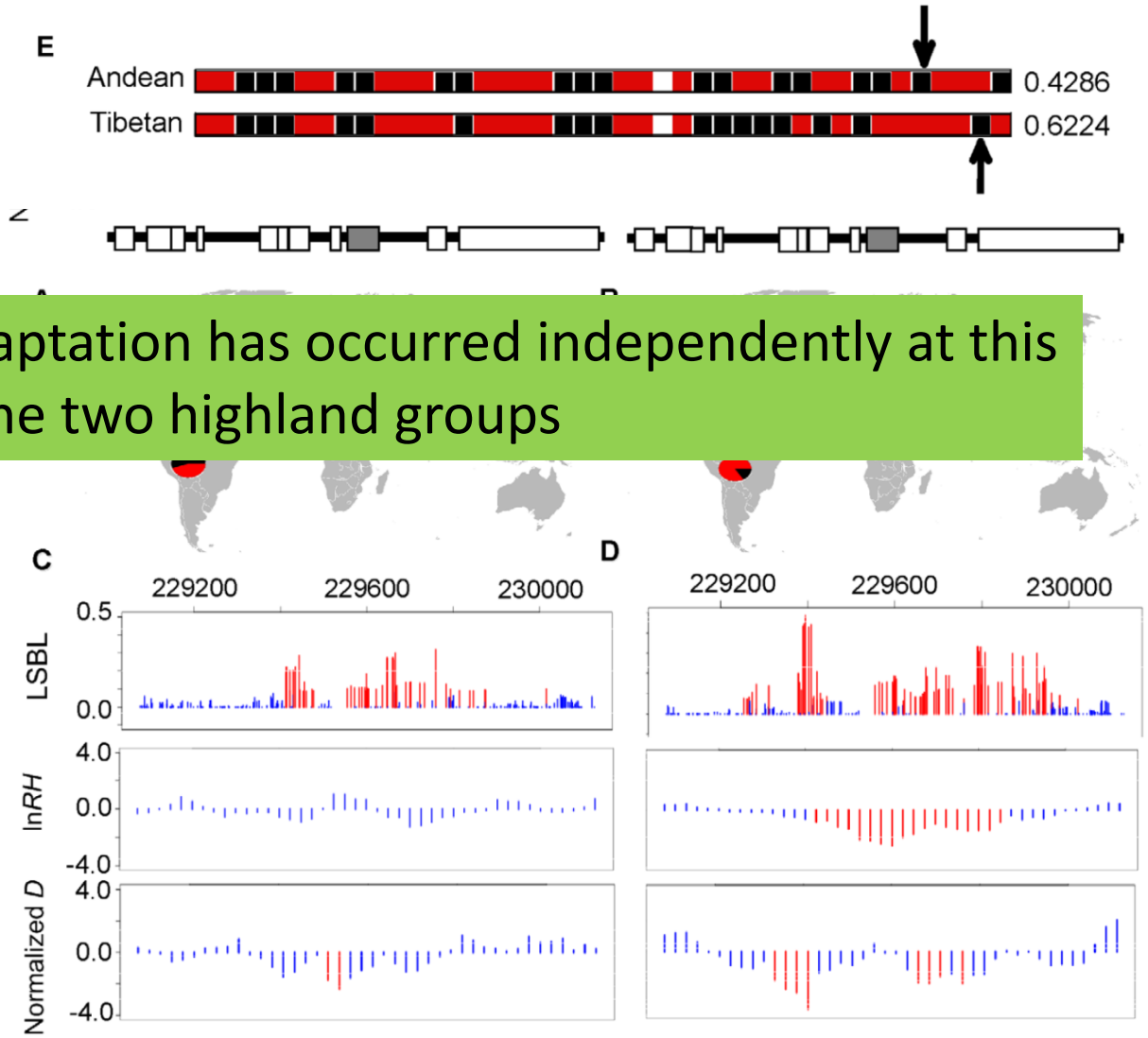
E: Haplotypes with arrow showing highest significant SNP

Grey region is gene

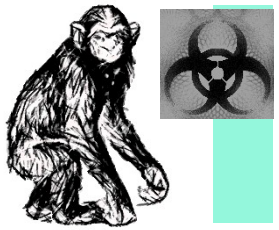
A&B: Allele frec
distribution of 2
ranked SNPs for Andeans and Tibetans
Derived =Red
Positive selection =Black

C: Significant are in Red for Andeans
D: and for Tibetans

THM: Adaptation has occurred independently at this gene in the two highland groups

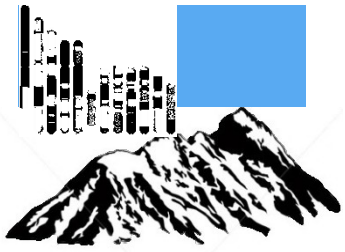


Take Home Message



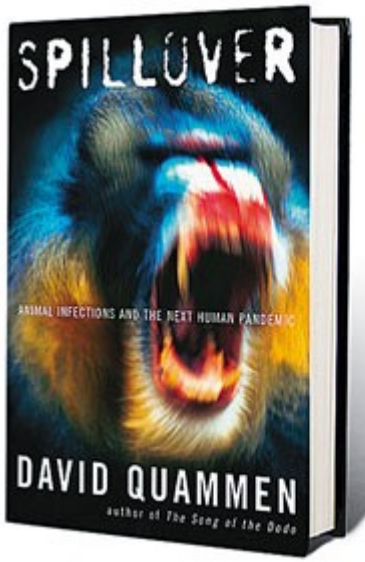
1. **The Chimp and the River**
   - Phylogenetic methods to detect selection in a parasite and host

2. **The Island Fox**
   - Balancing selection to resist effects of drift, but be careful with conclusions



3. **Men in the Mountains**
   - Positive selection across the genome can affect different region for convergent phenotypes

Acknowledgements



The excellent popular science book **Spillover: Animal Infections and the Next Human Pandemic** by David Quammen

Funding Sources:
European Social Fund in the Czech Republic, European Union, Ministry of Education, OP Education for Competitiveness, Veda vsemi smysly (CZ.1.07/2.3.00/35.0026)

INSTITUTE OF VERTEBRATE BIOLOGY ACADEMY OF SCIENCES CR

BOTANICKÝ ÚSTAV AV ČR v.v.i.

esf european social fund in the czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION, YOUTH AND SPORTS

OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

Thanks for your attention!