

# Recent selection in Tibet, Greenland & maybe China



Anders Albrechtsen

March 29, 2022

### Signatures of recent/ongoing selection

●○○○○

Tibet

○

A horizontal row of 18 small circles, evenly spaced, representing the total number of circles in the pattern.

## Greenland

A horizontal row of 15 small circles, each with a thin black outline.

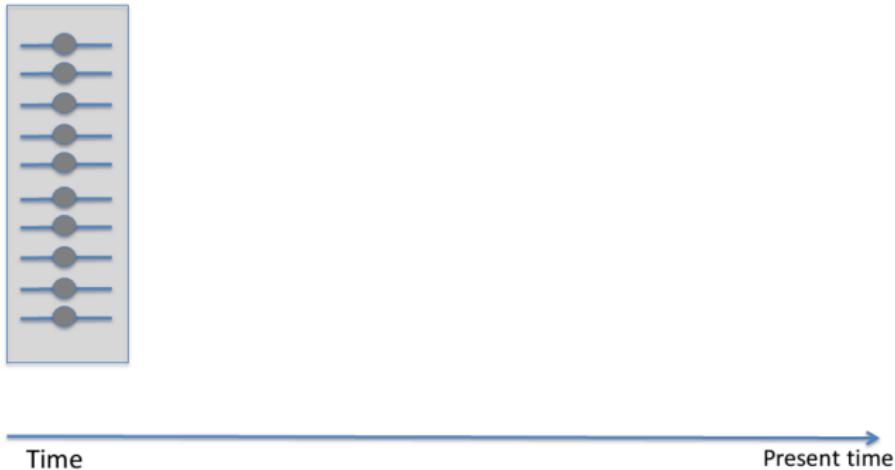
SFS for NGS data

6

6

○○○○○○○

## Allele frequency differentiation



## Signatures of recent/ongoing selection

○●○○○

Tibet

○

## Greenland

C

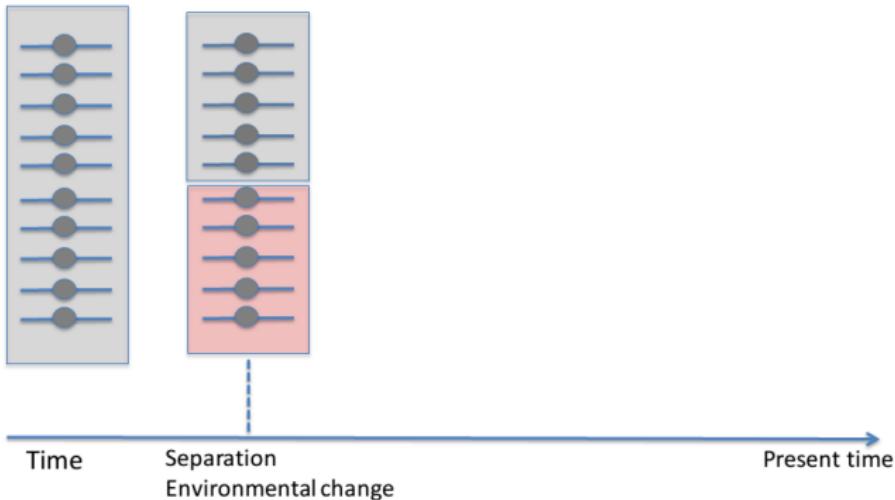
SFS for NGS data

C

C

○○○○○○○

## Allele frequency differentiation



## Signatures of recent/ongoing selection

○○●○○

## Tibet

A sequence of 15 small circles arranged in two rows. One circle is positioned in the top row, and 14 circles are in the bottom row, forming a horizontal line.

## Greenland

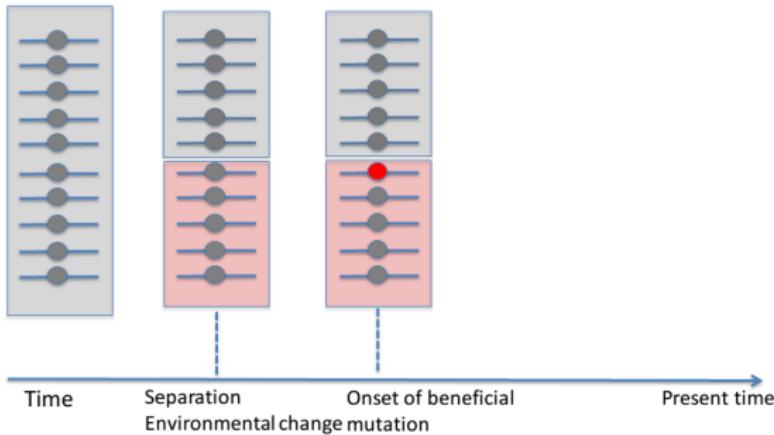
A horizontal row of 15 small, light brown circles arranged in a single line.

SFS for NGS data

1

# Probability of fixation

## Allele frequency differentiation



## Signatures of recent/ongoing selection

○○○●○

Tibet

A sequence of 15 small circles arranged in two rows. One circle is positioned in the top row, and 14 circles are in the bottom row, creating a total of 15 circles.

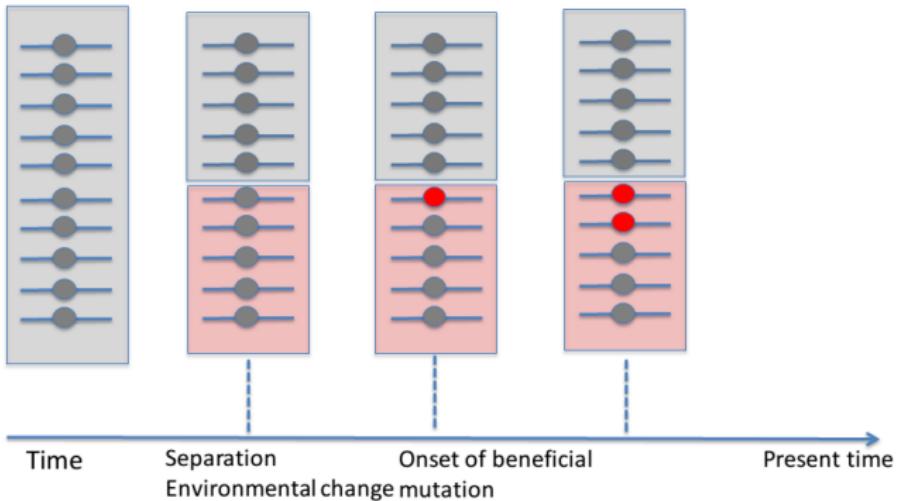
## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a visual element in the document.

SFS for NGS data

10

## Allele frequency differentiation



## Signatures of recent/ongoing selection

○○○○●

Tibet

A sequence of 15 small circles arranged in two rows. One circle is positioned in the top row, and 14 circles are in the bottom row, creating a total of 15 circles.

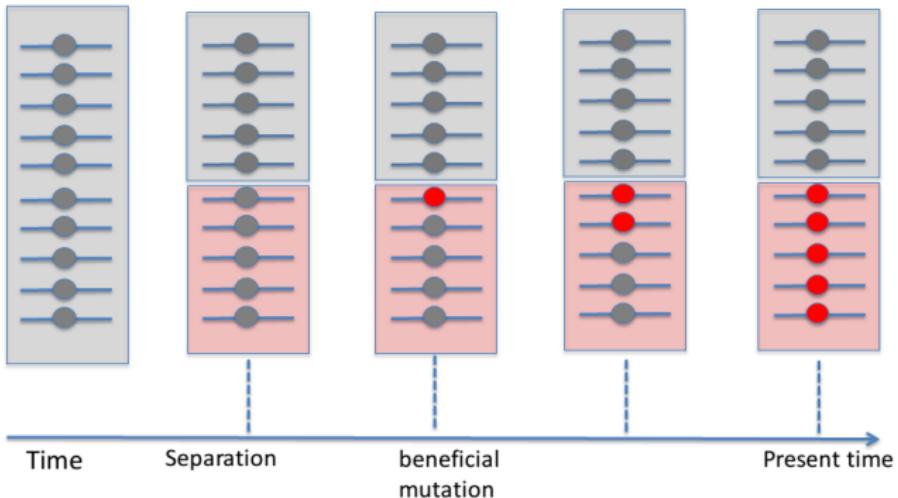
## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a visual element.

## SFS for NGS data

10

## Allele frequency differentiation



## Signatures of recent/ongoing selection

○○○○○

## Tibet

## Greenland

A horizontal row of 15 small, light gray circles, evenly spaced from left to right.

SFS for NGS data

10

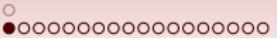
# Adaptation in Tibet



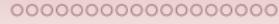
Photo by Crystal  
©2007 Crystal Main

### Signatures of recent/ongoing selection

## Tibet



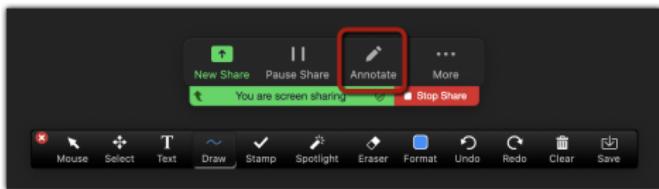
## Greenland



SFS for NGS data



## What are the possible drivers of selection for Inuit?



## Figure: Use the Text

### Signatures of recent/ongoing selection

## Tibet

○

A horizontal sequence of 18 circles. The first circle is black, and the remaining 17 are white.

## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a visual element in the document.

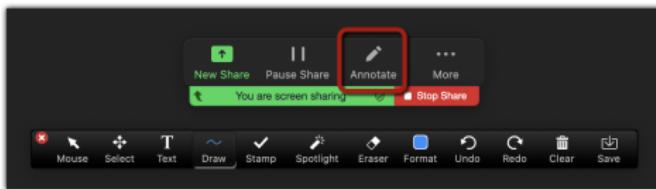
## SFS for NGS data

1

4

○○○○○○○

## What phenotypes might be affected?



## Figure: Use the Text

## Signatures of recent/ongoing selection

○○○○○

## Tibet

A horizontal sequence of 18 small circles. The first circle is light blue, the second is dark blue, and the remaining 16 are light blue.

## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a decorative element.

SFS for NGS data

10

## Altitude adaption in Tibet

Yi et al. 2010

- Low oxygen has a large effect on fitness
  - People living in high altitude generally have more birth defects

Photo by Crystal  
©2007 Crystal Main

## Signatures of recent/ongoing selection

○○○○○

## Tibet

○

Greenland

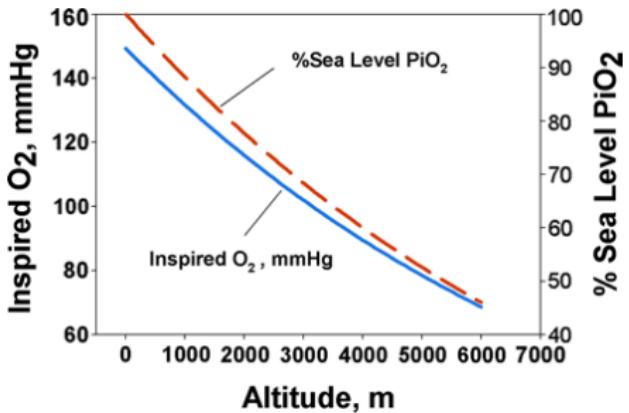
A horizontal row of 15 small circles, evenly spaced, used as a visual element.

SFS for NGS data

C

○

## Oxygen and height



### Signatures of recent/ongoing selection

## Tibet

○

A horizontal row of fifteen empty circles. The 6th circle from the left is filled black.

## Greenland

A horizontal row of 18 small circles, evenly spaced, used as a visual element in the page header.

SFS for NGS data

1

1

600

## Altitude adaption in Tibet

Yi et al. 2010

- The full exomes of 50 Tibetan individuals at an average coverage of 18X.
  - Compared to 40 Han Chinese individuals sequenced at an average of 6X (1000G).
  - Estimated joint allele frequencies for each SNP using Bayesian approach.

Signatures of recent/ongoing selection  
ooooo

Tibet

○

Greenland

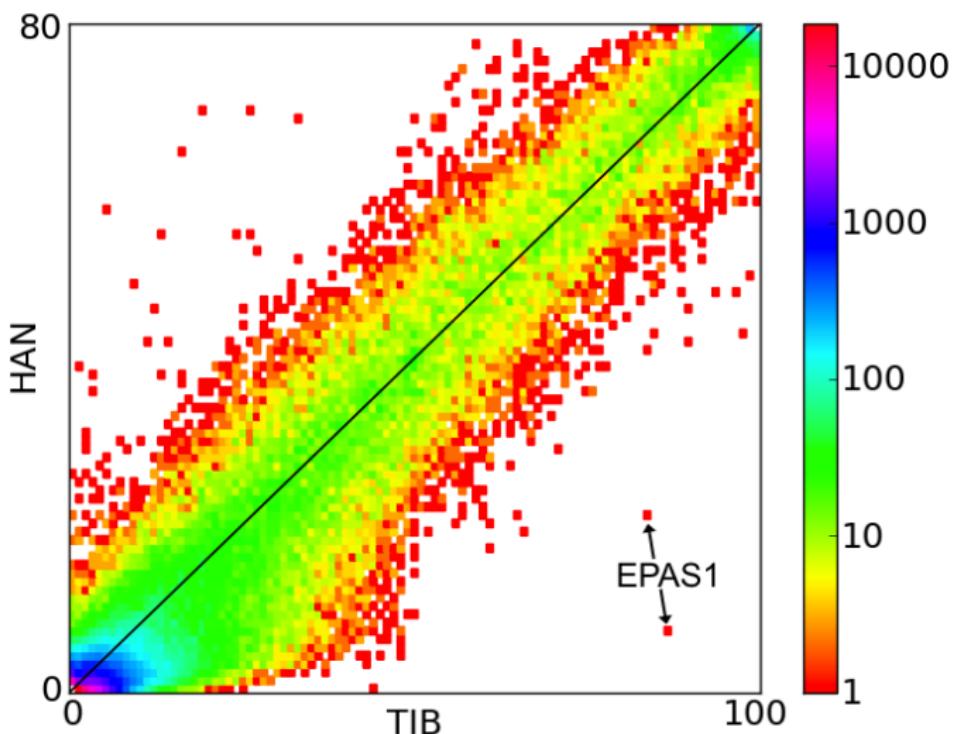
A horizontal row of 15 small, light-colored circles arranged in a single line.

SFS for NGS data

C

○

## 2D site frequency spectrum



## Signatures of recent/ongoing selection

○○○○○

## Tibet

○

A horizontal sequence of 18 small circles. The 10th circle from the left is filled black, while all other circles are unfilled.

Greenland

A horizontal row of fifteen small, light gray circles, evenly spaced from left to right.

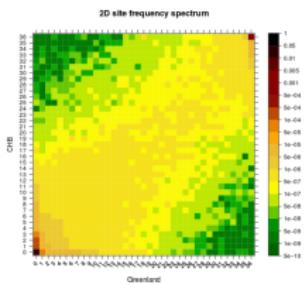
SFS for NGS data

C

C

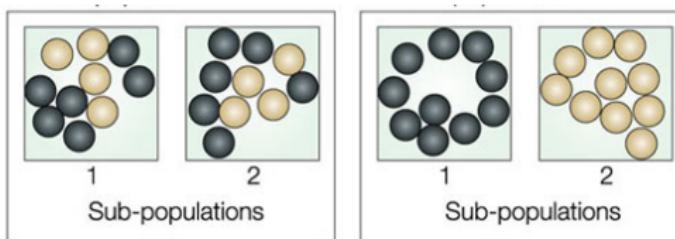
○○○○○○○

## 2D SFS and Fst



## Fst from heterozygosity

$$F_{st} = \frac{\sigma_B}{\sigma_T} = \frac{H_{total} - H_{subpopulations}}{H_{total}}$$



$$F_{ST} = 0$$

## Signatures of recent/ongoing selection

## Tibet

## Greenland

SFS for NGS data

1

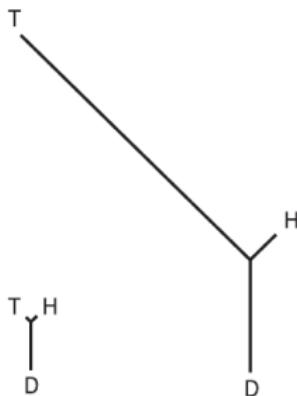
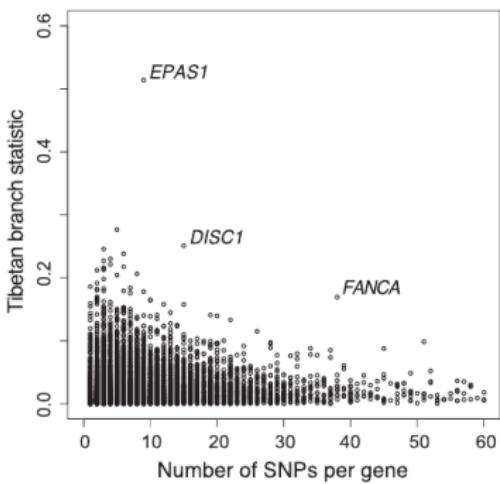


A horizontal row of 15 small circles, evenly spaced, used as a decorative element.

○

# Population Branch Statistic (PBS)

$$PBS = TBS = (T^{TH} + T^{TD} - T^{HD})/2, \quad T^{AB} = -\log(1 - F_{st}^{AB})$$

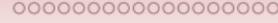


Signatures of recent/ongoing selection  
ooooo

## Tibet



## Greenland



SFS for NGS data



# Population frequencies

## EPAS1 SNP allele frequencies

Allele	Tibetan	Han	Danish
C	0.13	0.9125	1
G	0.87	0.0875	0

## Signatures of recent/ongoing selection

○○○○○

# Tibet

○

## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a decorative element.

SFS for NGS data

6

6

○○○○○○○

EPAS1

- type of hypoxia-inducible factors
  - active under low oxygen
  - variant of gene confers increased athletic performance - called the "super athlete gene".

## Signatures of recent/ongoing selection

## Tibet

○

A horizontal sequence of 15 small circles. The 13th circle from the left is filled black, while all other circles are unfilled.

## Greenland

A horizontal row of fifteen empty circles, intended for a child to connect with straight lines to form a continuous path.

SFS for NGS data

C

6

○○○○○○○

## Genotyping in 366 individuals

## Independent genotyping

- 366 Tibetans
  - Genotyped for the EPAS1 SNP
  - Phenotypes available

## Associations within the Tibetan population

	CC	CG	GG	p-value
N	10	84	272	
Hemoglobin concentration	178	178.9	167.5	0.0013
erythrocyte counts	5.3	5.6	5.2	0.0015

## Signatures of recent/ongoing selection

## Tibet

## Greenland

SFS for NGS data

○○○○○

A horizontal row of 15 small, uniform circles arranged in a single line.

1

**Is this extreme compared to populations**



## Africans

- 1 Bantu
  - 2 Mandenka
  - 3 Yoruba
  - 4 San
  - 5 Mbuti pygmy
  - 6 Biaka
  - 7 Mozabito

Europeans

- 8 Orcadian
  - 9 Adygei
  - 10 Russian
  - 11 Basque
  - 12 French
  - 13 North Italian
  - 14 Sardinian
  - 15 Tuscan

## Western Asians

- 16 Bedouin
  - 17 Druze
  - 18 Palestinian

## Central and Southern Asians

- |            |             |
|------------|-------------|
| 19 Balochi | 35 Orogen   |
| 20 Brahui  | 36 She      |
| 21 Makrani | 37 Tujia    |
| 22 Sindhi  | 38 Tu       |
| 23 Pathan  | 39 Xibo     |
| 24 Burusho | 40 Yi       |
| 25 Hazara  | 41 Mongolia |

## Eastern Asians

- 28 Han (S. China)
  - 29 Han (N. China)
  - 30 Dai
  - 31 Daur
  - 32 Hezhen

32 Hsieh  
33 Lahu  
34 Miao

- 34 Miao
  - 35 Oroqen
  - 36 She
  - 37 Tujia
  - 38 Tu
  - 39 Xibo
  - 40 Yi
  - 41 Mongola

### Oceanians

- 46 Melanesian  
47 Papuan

## Native Americans

- 48 Karitiana
  - 49 Surui
  - 50 Colombian
  - 51 Maya
  - 52 Pima

## Signatures of recent/ongoing selection

## Tibet

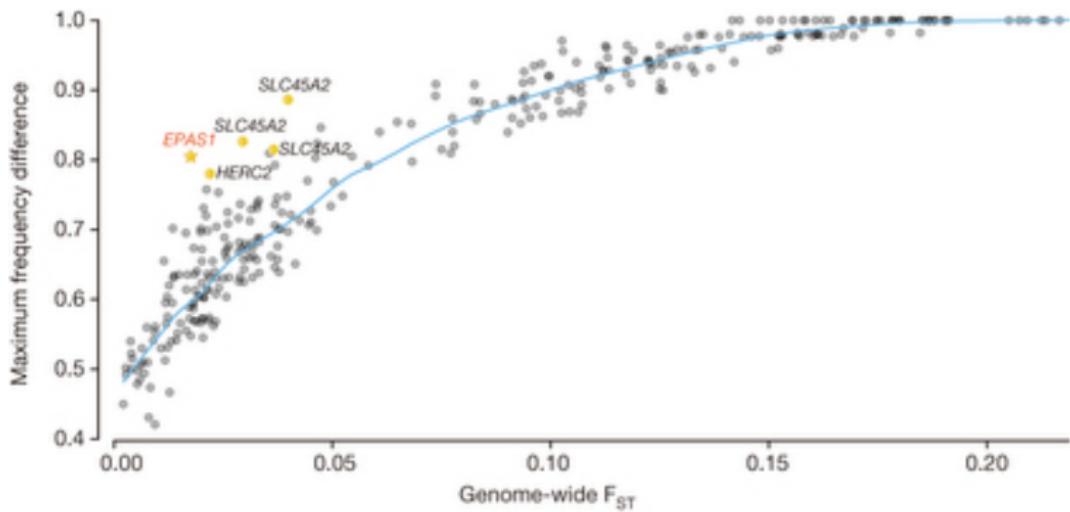
## Greenland

## SFS for NGS data

○○○○○

A horizontal row of fifteen small, light gray circles, evenly spaced from left to right.

## Other genes with large FST



Signatures of recent/ongoing selection

○○○○○

Tibet

○  
○○○○○○○○○○○○○○●○○○○

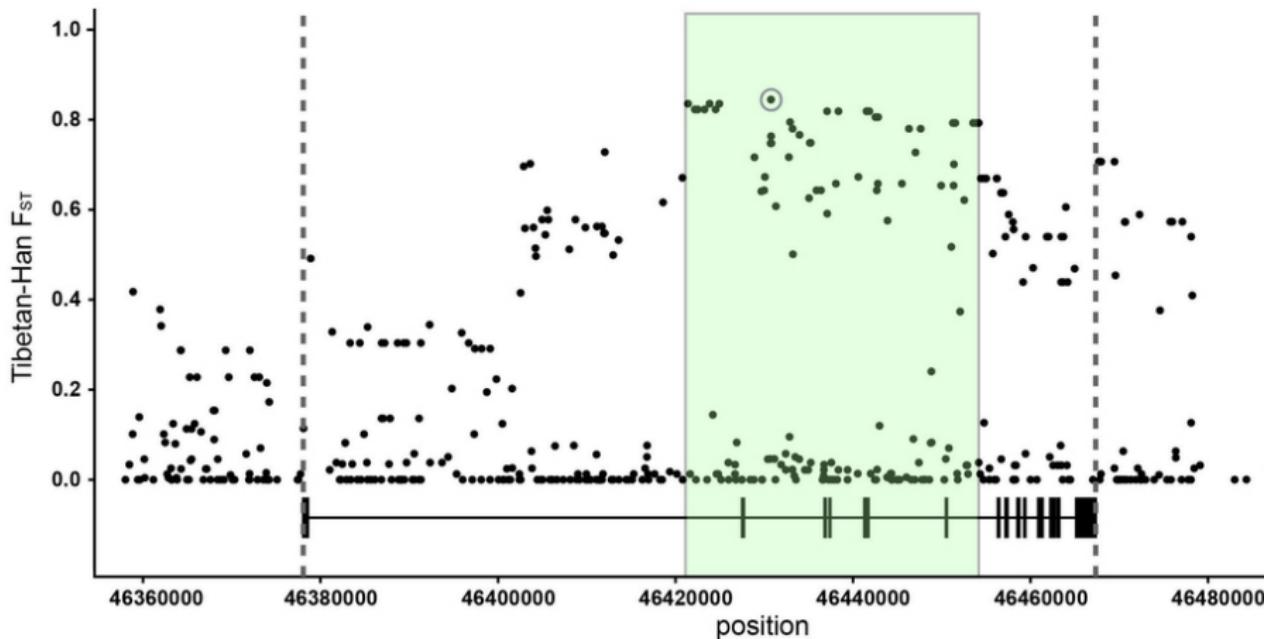
Greenland

○○○○○○○○○○○○○○○○○○○○○○

SFS for NGS data

○  
○  
○○○○○○

## EPAS1



Signatures of recent/ongoing selection

oooooo

Tibet

○  
oooooooooooooooooooo●oooo

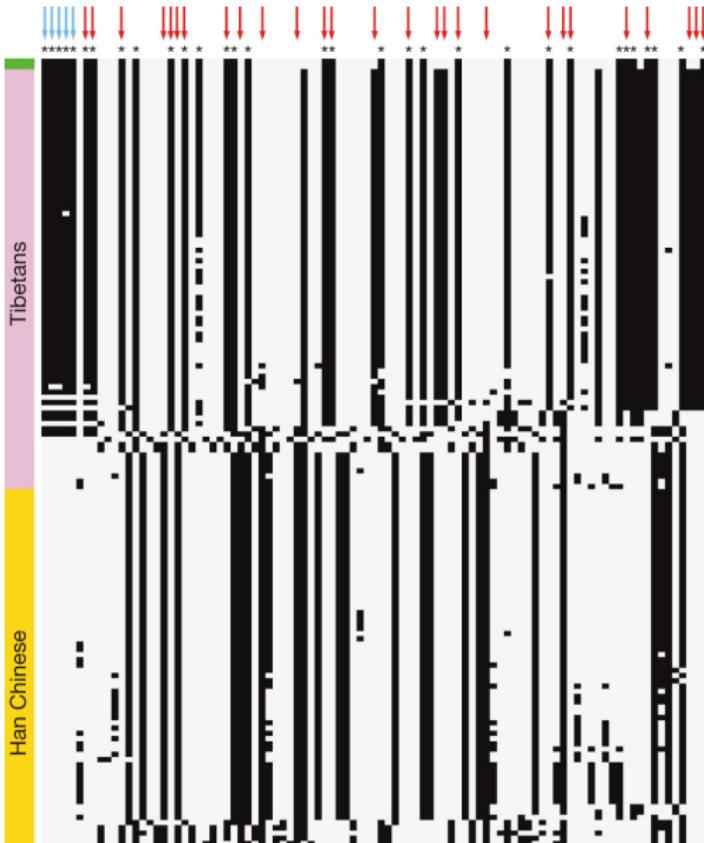
Greenland

ooooooooooooooooooooooo

SFS for NGS data

○  
○  
ooooooo

## Haplotype is extremely different

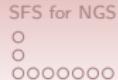


### Signatures of recent/ongoing selection

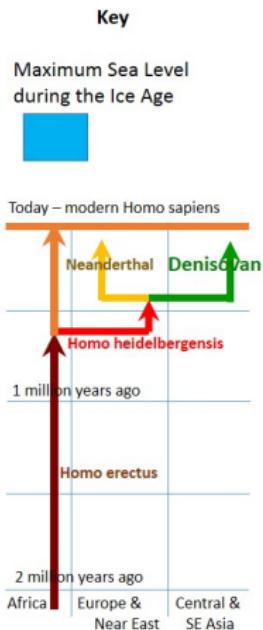
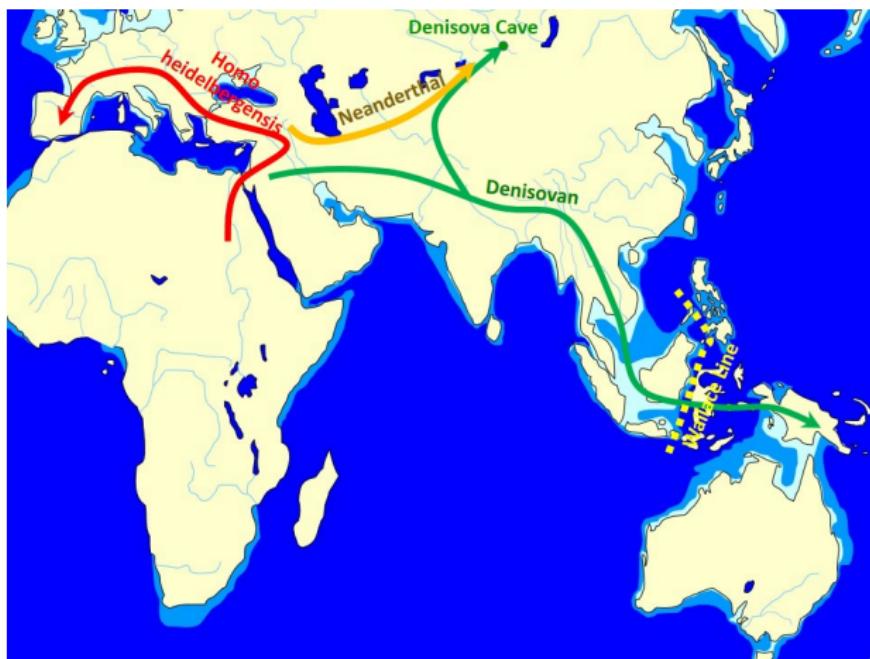
## Tibet

## Greenland

SFS for NGS data



## How did they adapt so fast



Signatures of recent/ongoing selection

oooooo

Tibet

○  
oooooooooooooooooooo●○

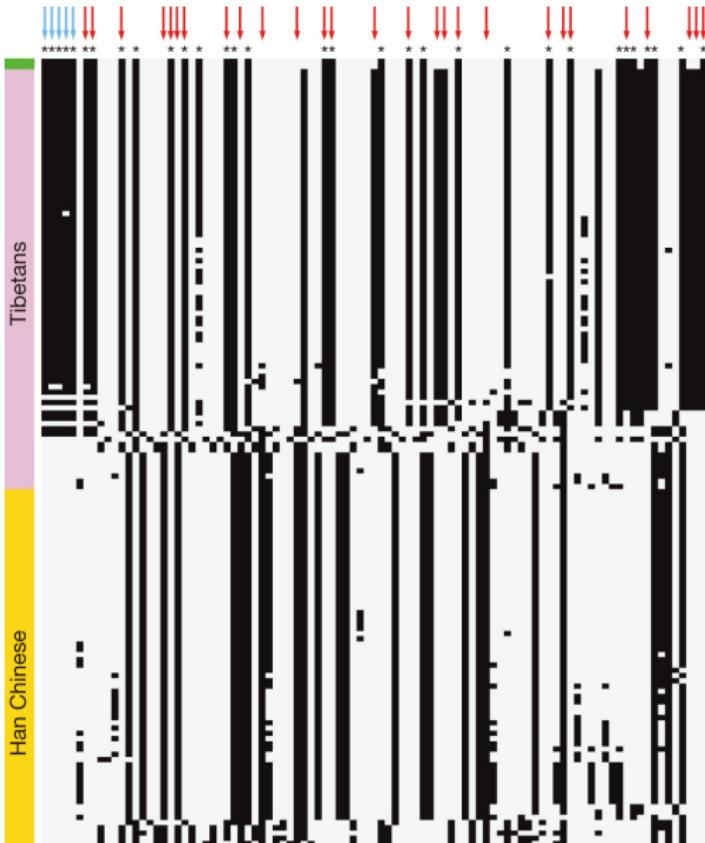
Greenland

ooooooooooooooooooooooo

SFS for NGS data

○  
○  
ooooooo

## Adaptive intergression



## Signatures of recent/ongoing selection

○○○○○

## Tibet

○

○○○○○○○○○○○○○○○●

## Greenland

A horizontal row of 15 small circles, each containing a dot, used as a decorative element.

SFS for NGS data

C

6

886

## Conclusion

- Tibetans have adapted to life in high altitude
  - A loci EPAS1 was found that has undergone strong adaptive selection
  - The loci associated with hemoglobin concentrations and erythrocyte counts
  - The mutations were introduced by Denisovan introgression
  - First (and only) example of adaptive introgression in humans

### Signatures of recent/ongoing selection

Tibet

Greenland

SFS for NGS data

○○○○○

A horizontal sequence of 15 circles. The first circle is solid black, and the subsequent 14 circles are hollow with a dark outline.

○

## **Human adaption to arctic environment**



Signatures of recent/ongoing selection

ooooo

Tibet

○  
oooooooooooooooooooo

Greenland

○●oooooooooooooooooooo

SFS for NGS data

○  
○  
ooooooo

## Brief overview of Greenland's history



- Inhabited on and off by different Arctic cultures for ~4500 years:
- 
- Visited by Vikings, Danish colony from 1814, now autonomous country



## Signatures of recent/ongoing selection

## Tibet

## Greenland

SFS for NGS data

○○○○○

A horizontal row of fifteen empty circles. The 6th circle from the left is filled black, while all others are empty.

20

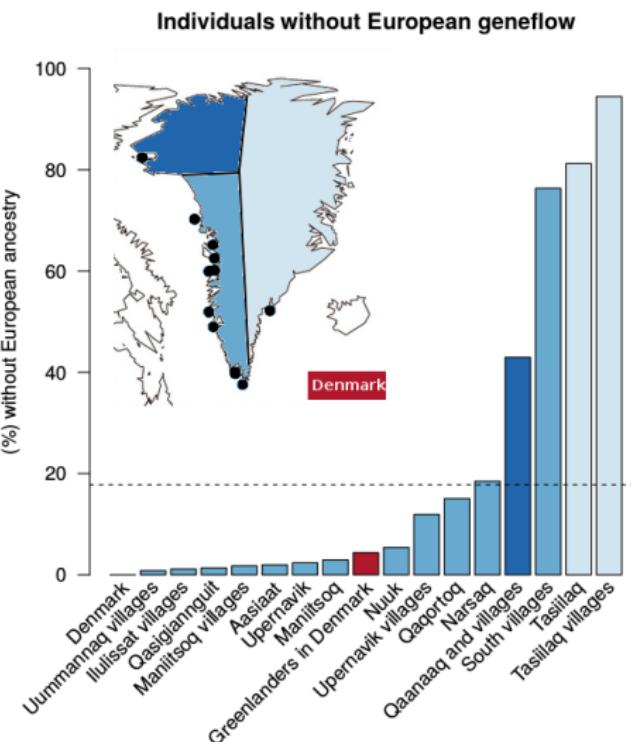
## The modern Greenlandic population

- Small:  $N \approx 57,000$
  - Live in coastal towns
  - Descendents of Inuit



- But most also have European ancestry
  - On average ~ 25%

From Moltke et al. 2014, AJHG



## Signatures of recent/ongoing selection

Tibet

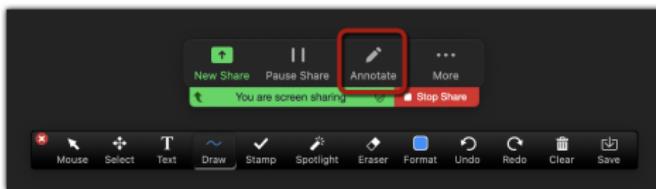
## Greenland

SFS for NGS data

○○○○○

10

## What are the possible drivers of selection for Inuit?



## Figure: Use the Text

## Signatures of recent/ongoing selection

## Tibet

Greenland

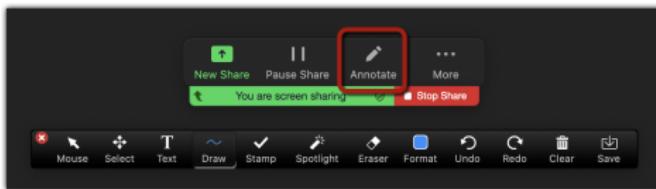
SFS for NGS data

○○○○

A horizontal sequence of 15 small circles, with one circle at the far left and 14 circles following it.

○

## What phenotypes might be affected?



## Figure: Use the Text

Signatures of recent/ongoing selection

oooooo

Tibet

o  
oooooooooooooooooooo

Greenland

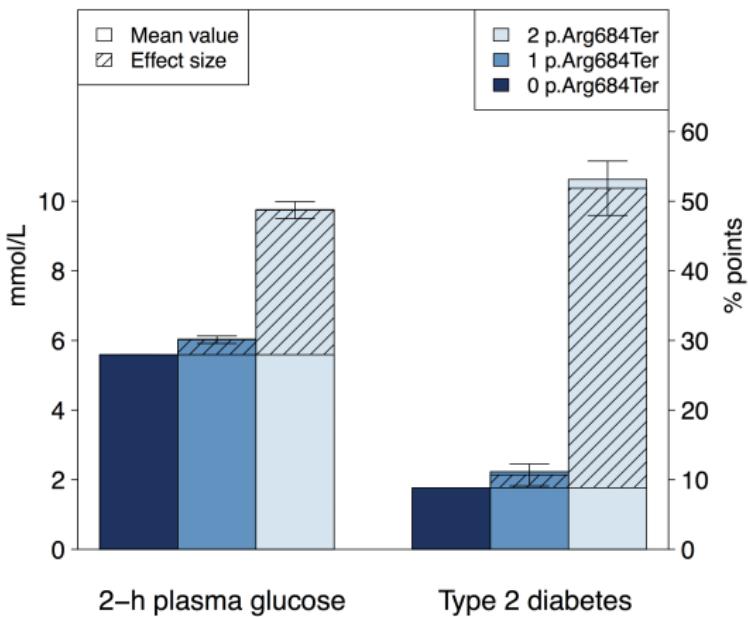
oooooo●oooooooooooo

SFS for NGS data

o  
o  
ooooooo

## A mutation causes 15% of type 2 diabetes in Greenland<sup>1</sup>

Very large almost recessive effect



Rec model

2-h glucose T2D:  
OR=10.3

heredibility

The variation explain  
15% of all T2D in  
Greenland

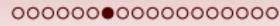
<sup>1</sup>Moltke et al. Nature, 2014

### Signatures of recent/ongoing selection

Tibet



## Greenland



SFS for NGS data



**Life in the Arctic is extreme: cold temperatures & fat-rich diet**



### Signatures of recent/ongoing selection

## Tibet



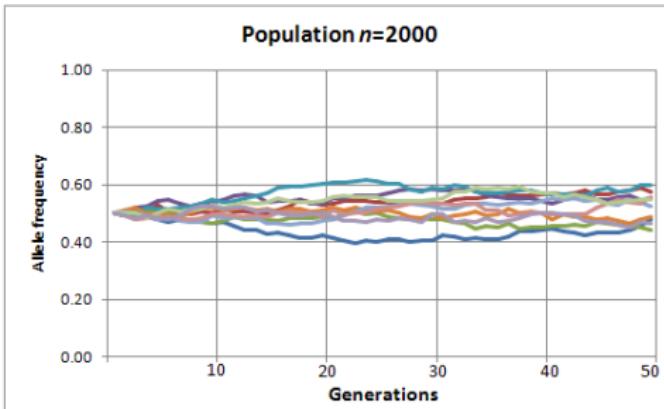
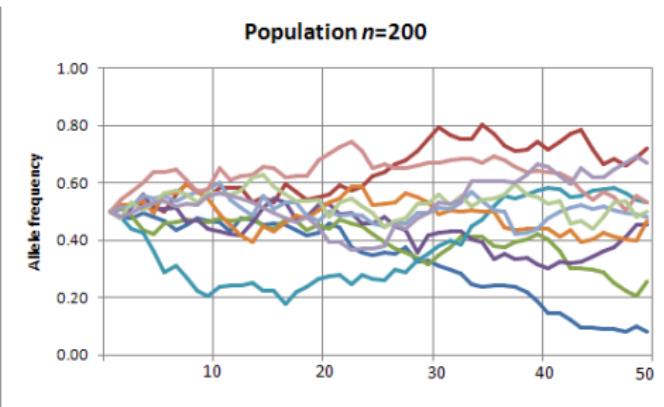
## Greenland



SFS for NGS data



## Allele frequencies and population size



Signatures of recent/ongoing selection

oooooo

Tibet

o  
oooooooooooooooooooo

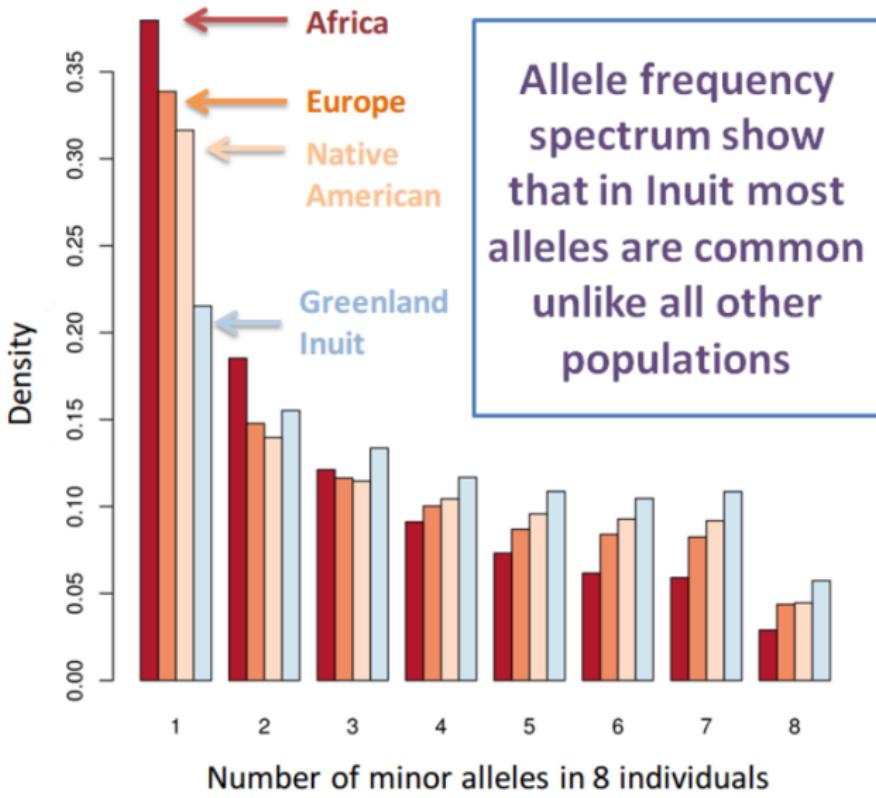
Greenland

oooooooooooo●oooooooooooo

SFS for NGS data

o  
o  
ooooooo

## Frequency spectrum of Inuit



## Signatures of recent/ongoing selection

## Tibet

Greenland

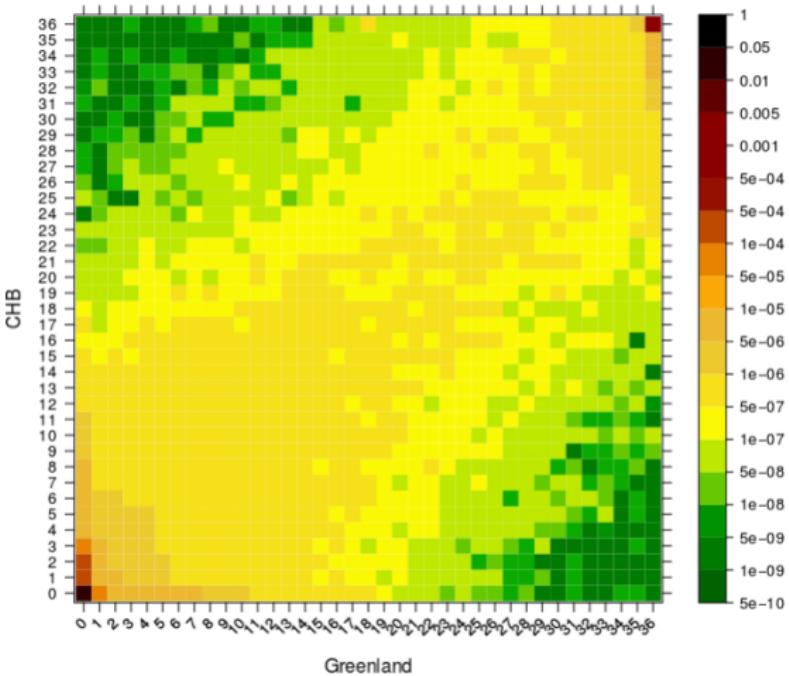
## SFS for NGS data

—

10

## 2D SFS between GL and Han

## 2D site frequency spectrum



### Signatures of recent/ongoing selection

## Tibet

## Greenland

SFS for NGS data

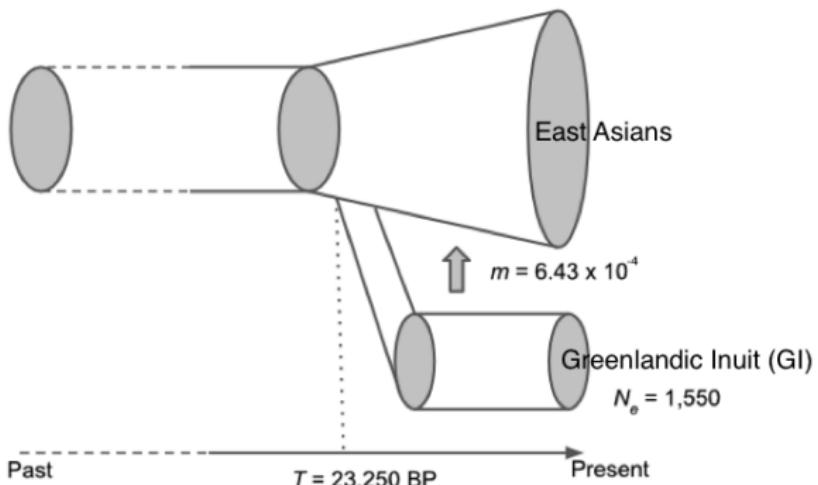
○○○○○

A horizontal row of 18 small circles. The 10th circle from the left is filled black, while all other circles are unfilled.

○

## Split time from East Asia

## Analyses of the exome data using *daði*:



## Tree based on Fst

## Genome-wide avg.

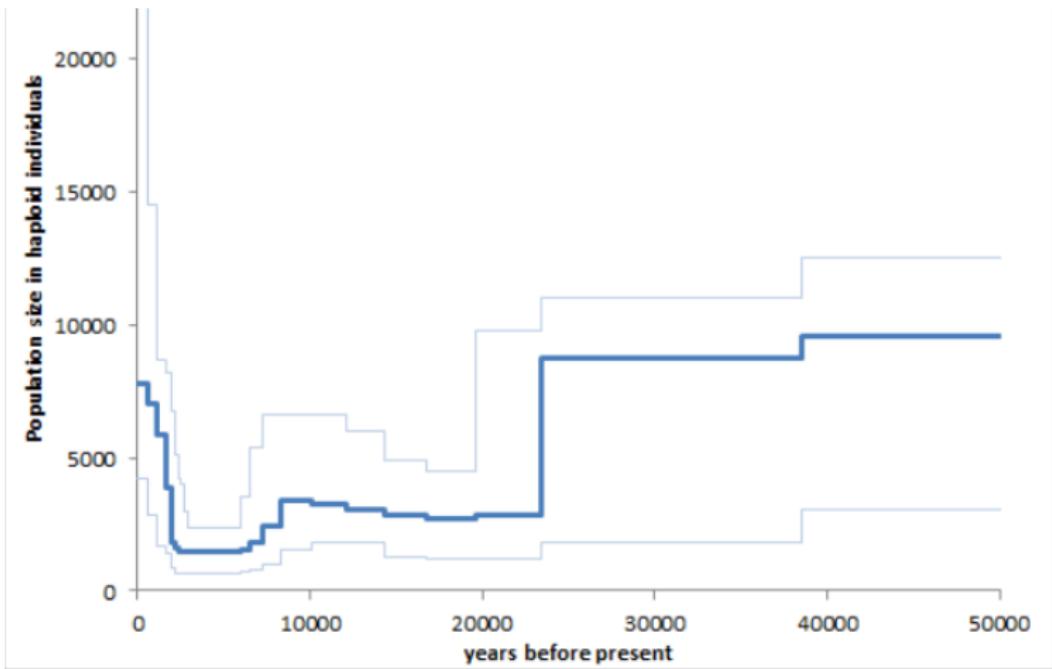
GI Han  
CEU

### Signatures of recent/ongoing selection

Tibet

Greenland

SFS for NGS data



## Signatures of recent/ongoing selection

## Tibet

Greenland

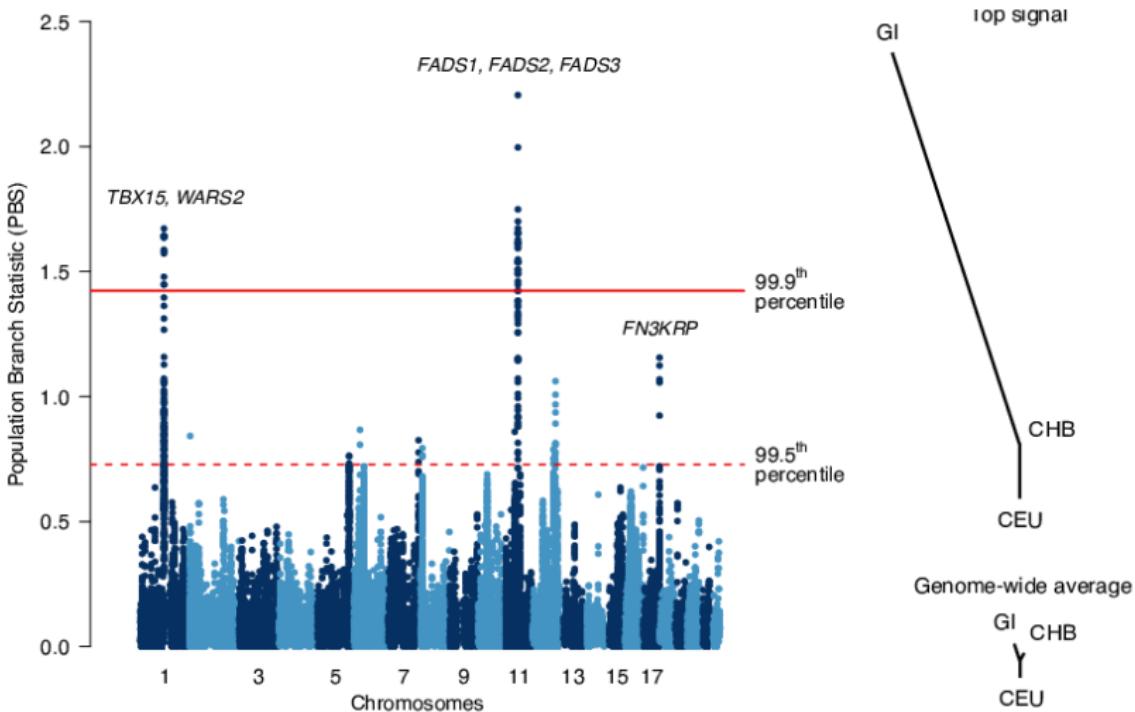
SFS for NGS data

○○○○○

A horizontal sequence of 15 small circles. The first 12 circles are hollow with a thin black outline, while the 13th circle from the left is filled with black color.

1

## Selection scan using PBS - ((HAN, GR) CEU)



Signatures of recent/ongoing selection

ooooo

Tibet

o  
oooooooooooooooooooo

Greenland

oooooooooooooooooooo●oooo

SFS for NGS data

o  
o  
ooooooo

## Top loci

### FADS

fatty acid desaturase.

### TBX15

- TBX15 plays an important role in differentiation of brown (subcutaneous) adipocytes.
- Upon stimulation by cold exposure can produce heat by lipid oxidation.

### FN3KRP

- an enzyme that catalyzes fructosamines, psicosamines and ribulosamines that protects against nonenzymatic glycation.
- FN3KRP can act to counteract the negative fitness caused by a PUFA rich diet.

## Signatures of recent/ongoing selection

## Tibet

Greenland

SFS for NGS data

○○○○○

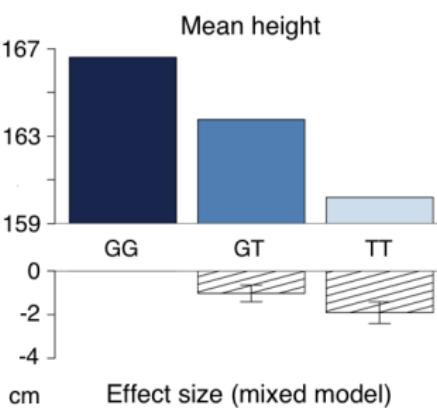
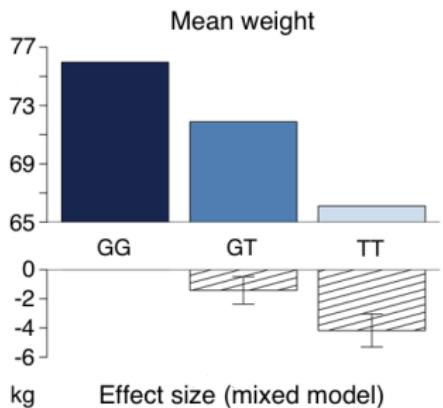
○ 88888888888888888888

○○○○○○○○○○○●○○○

○

## Why selection?

- Tested for association between top SNPs and metabolic traits
  - Marginally significant associations with multiple traits, including LDL
  - Selected alleles associated with decreased weight and height:



## Signatures of recent/ongoing selection

## Tibet

○

A horizontal row of 20 small circles, evenly spaced, representing the total number of circles available.

Greenland

A horizontal sequence of 15 circles. The first 14 circles are white with black outlines, arranged in a single row. The 15th circle, located at the end of the sequence, is solid black with no outline.

## SFS for NGS data

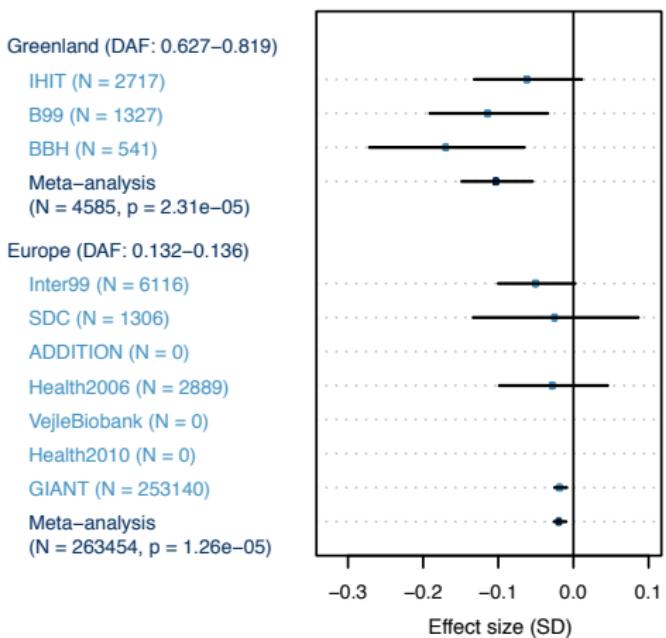
C

6

○○○○○○○

## Why selection?

- The association with height replicates in Europe:



## Signatures of recent/ongoing selection

## Tibet

Greenland

SFS for NGS data

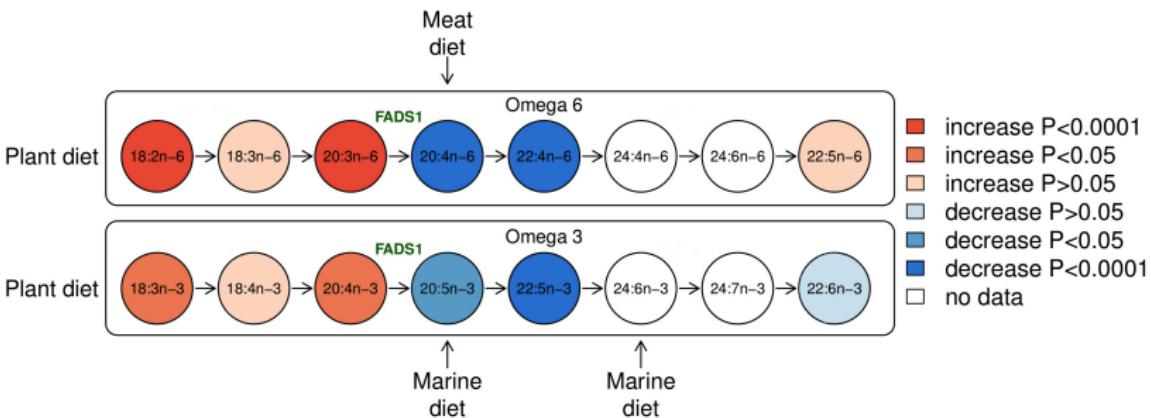
○○○○○

A horizontal sequence of 15 small circles. The first 14 circles are hollow with a thin black outline, while the 15th circle from the left is filled with black.

○

## Why selection? Take 2

- Testing for association w. red blood cell membrane fatty acid composition:



- Mutation seems to compensate for high-fat diet
  - Height due to effect of fatty acid composition on growth hormone levels?
  - Either way, the results suggest that selection in this region is a new example of human adaptation where we know the genetic basis

Signatures of recent/ongoing selection

ooooo

Tibet

○  
oooooooooooooooooooo

Greenland

oooooooooooooooooooo●

SFS for NGS data

○  
○  
ooooooo

## Conclusion

- We find multiple loci with recent adaptation to life in the arctic
- As expected the genes are involved in poly unsaturated fatty acid metabolism and cold adaption
- Surprisingly the loci also affects hair and weight
- Mutations also have an effect in Europe

## Signatures of recent/ongoing selection

○○○○○

Tibet

Q

A horizontal row of 15 small circles, evenly spaced, used as a decorative element.

## Greenland

○○○○○○○○○○○○○○○○

SFS for NGS data

6

○  
○○○○○○

## How are the SFS estimated?

## Can we construct the SFS using NGS data

Yes - but be careful

## Signatures of recent/ongoing selection

## Tibet

## Greenland

SFS for NGS data

○○○○○

A horizontal sequence of 15 small circles. The first circle is positioned above the first circle in the sequence, creating a small gap at the start.

A horizontal row of 15 small, light gray circles, evenly spaced from left to right.

1

## When can calling SNPs and genotypes be a problem?

## low/medium depth data

- Capture data
  - low depth sequencing due to price
  - ancient DNA (only a finite amount of DNA)

What depth is high enough?

Depends on the analysis

- SFS is extremely sensitive to both genotype and SNP calling
  - admixture proportions are sensitive to genotype calling
  - ABBA-BABA (D-stats) can be used regardless of depth

## Signatures of recent/ongoing selection

## Tibet

○

A horizontal row of 15 small circles, each with a thin black outline.

## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a visual element in the document.

## SFS for NGS data

6

6

●○○○○○○

## Estimating SFS using uncertainty of the data

## Likelihood of SFS for a single site:

$$P(X^s \mid \eta) = \sum_{j=0}^{2N} p(X^s \mid \eta, J=j) p(J=j \mid \eta)$$

$$\propto \sum_{j=0}^{2N} \eta_j \sum_{g \in \{0,1,2\}^N} p(G=g \mid J=j) \prod_{i=1}^N P(X_i^s \mid G_i=g_i),$$

$$p(G = g \mid J = j)$$

$$p(G = g \mid J = j) = \binom{2N}{j} 2^{\sum_i^N l_1(g_i)}$$

when  $\sum_{i=1}^{2N} g_i = j$ , else 0

## SFS for a region

$$P(X \mid \eta) = \prod_{s=1}^r P(X^s \mid \eta)$$

fast calculations with dynamic programming and EM<sup>2</sup>

<sup>2</sup>Nielsen et al. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data

### Signatures of recent/ongoing selection

○○○○○

## Tibet

A sequence of 15 small circles arranged in two rows. One circle is positioned in the top row, and 14 circles are in the bottom row, forming a horizontal line.

## Greenland

A horizontal row of 15 small circles, evenly spaced, used as a visual element in the document.

SFS for NGS data

1

## SFS based on genotype likelihoods

- can be estimate even with low(ish) depth e.g. 2 X
  - Must be done with genotype likelihoods unless depth is high ( $>10X$ )
  - Can be done in any dimension
    - 1D  $\theta$ s e.g. Tajimas pi, Tajimas D, Population sizes
    - 2D  $f_{st}$  and PBS
    - >2D useful for Demography inference

### Signatures of recent/ongoing selection

## Tibet

○

A horizontal row of 15 small circles, each containing a dot, used as a visual representation for the number fifteen.

Greenland

A horizontal row of 15 small circles, each with a thin gray outline.

## SFS for NGS data

C

C

○○●○○○○

## Time for exercises

## Data from 1000 Genomes

- 2500 individuals sequenced at low/medium depth (3-8X)
  - multiple populations

## Human genomes

- 3Gb
  - BAM file size 5Gb per X

## Reduced genome

- 22 100k regions (one for each autosome)
  - 1Mb region on chr5
  - 3 x 10 individuals from
  - African(YRI), European (CEU), East Asian (JPT)

## Signatures of recent/ongoing selection

## Tibet

## Greenland

SFS for NGS data

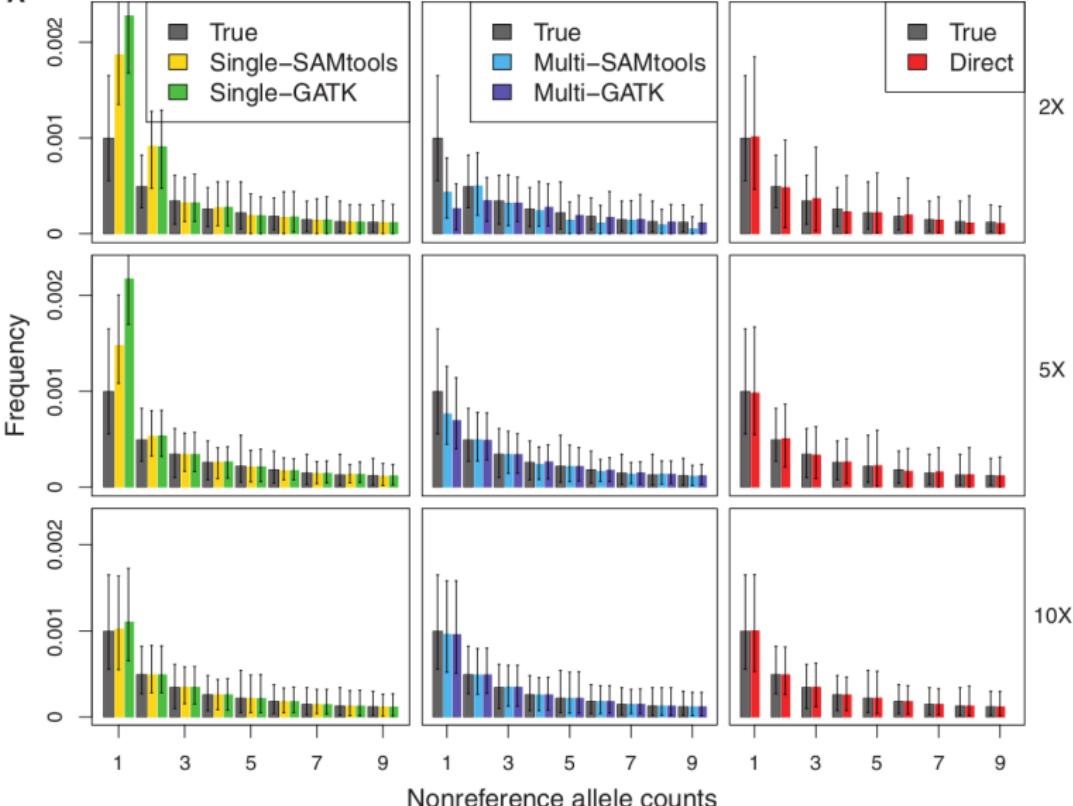
—

A horizontal row of 15 small, uniform circles arranged in a single line.

oo

## Site frequency spectrum for low/medium depth data

A



## Signatures of recent/ongoing selection

Tibet

## Greenland

SFS for NGS data

○○○○○

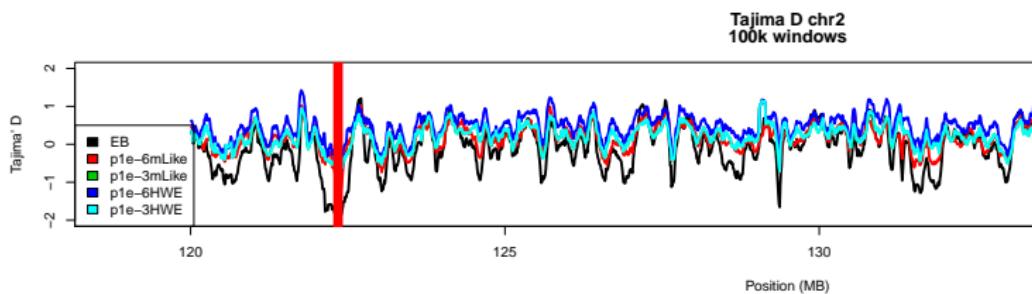
A horizontal sequence of 15 small circles. The first circle is positioned above the first circle in the main sequence. The main sequence consists of 14 circles arranged in a single row.

A horizontal row of 15 small circles, evenly spaced, used as a decorative element.

○

## Selection scan using emperical bayes

## LCT loci in Europeans based on 3X data



Signatures of recent/ongoing selection

oooooo

Tibet

o  
oooooooooooooooooooo

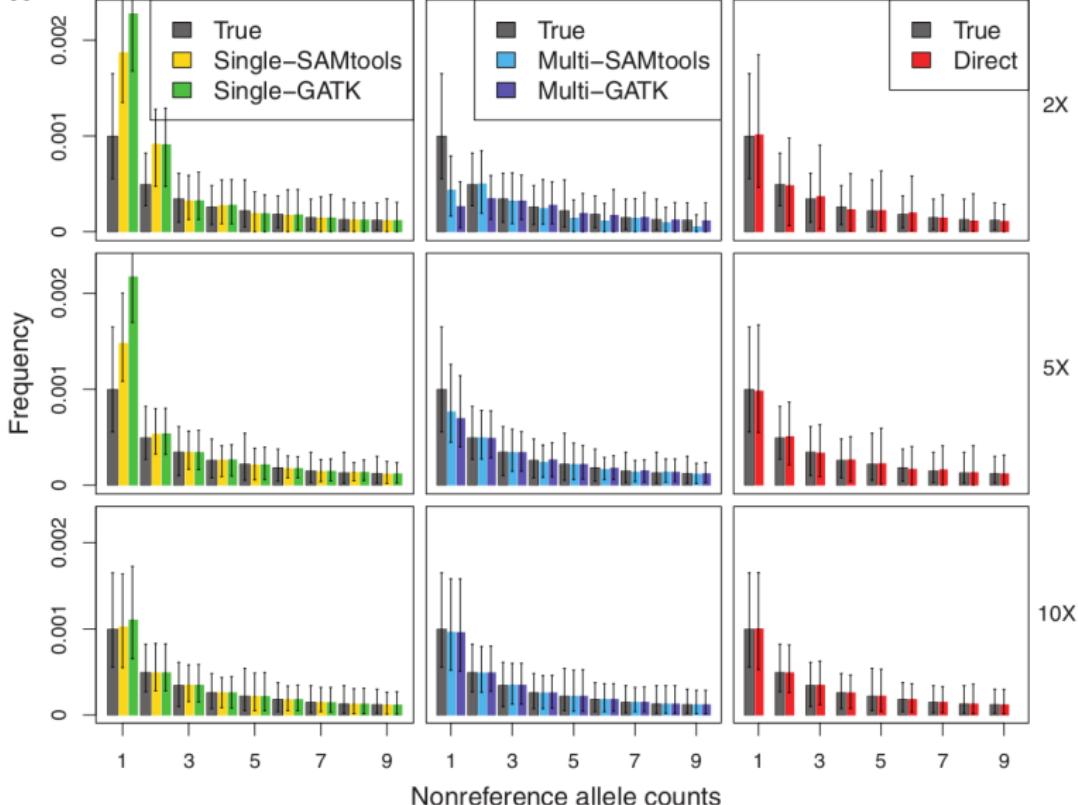
Greenland

oooooooooooooooooooo

SFS for NGS data

o  
o  
oooooo●○

## Site frequency spectrum for low/medium depth data

**A**

Signatures of recent/ongoing selection

oooooo

Tibet

○  
oooooooooooooooooooo

Greenland

oooooooooooooooooooo

SFS for NGS data

○  
○  
○○○○●

# There are no possible filters than can solve the problem

