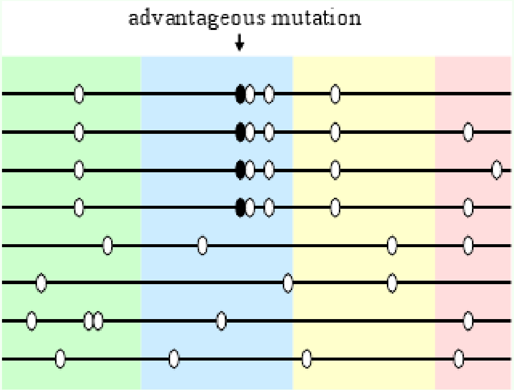
Selection and haplotypes
EHH statistics

Anders Albrechtsen

Signature of selection

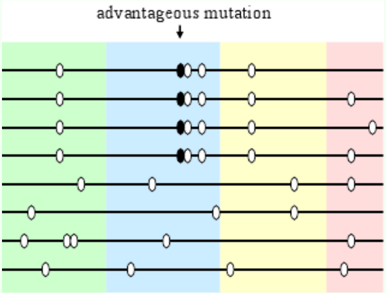
- Mutation enters the population
- Mutation increases in frequency due to positive selection
- **Increases LD**
- Affects the variability
- **Increases haplotype similarity**
- Increases differences with other populations in the whole region



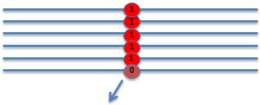
EHH - Extended Haplotype Homozygosity

What is EHH?

Extended haplotype homozygosity (EHH): EHH at distance x from the core region is the probability that two randomly chosen chromosomes carry a tested core haplotype are homozygous at all SNPs for the entire interval from the core region to the distance x .



Extended Haplotype Homozygosity



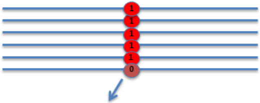
Core haplotype is 1
(Biallelic: 0 is ancestral, 1 is derived allele)

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

Core SNP

slides stolen from Matteo Fumagalli

Extended Haplotype Homozygosity

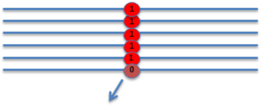


Core haplotype is 1
(Biallelic: 0 is ancestral, 1 is derived allele)

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

Until marker x_i
(starting from x_0)

Extended Haplotype Homozygosity

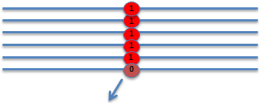


Core haplotype is 1
(Biallelic: 0 is ancestral, 1 is derived allele)

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

Sum across all unique haplotypes
carrying the core SNP

Extended Haplotype Homozygosity



Core haplotype is 1
(Biallelic: 0 is ancestral, 1 is derived allele)

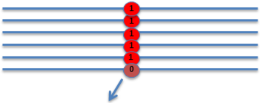
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

n_h is haplotype frequency of *h*

n_h is haplotype frequency of the core SNP

Sum across all unique haplotypes carrying the core SNP

Extended Haplotype Homozygosity



Core haplotype is 1
(Biallelic: 0 is ancestral, 1 is derived allele)

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

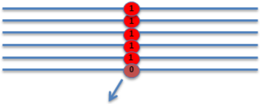
n_h is haplotype frequency of h

n_h is haplotype frequency of the core SNP

Sum across all unique haplotypes carrying the core SNP

$$EHH_c(x_i = 0) = ?$$

Extended Haplotype Homozygosity



Core haplotype is 1
(Biallelic: 0 is ancestral, 1 is derived allele)

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

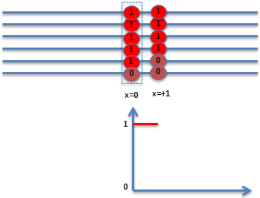
n_h is haplotype frequency of h

n_c is haplotype frequency of the core SNP

Sum across all unique haplotypes carrying the core SNP

$$EHH_c(x_i = 0) = \frac{\binom{5}{2}}{\binom{5}{2}} = 1$$

Extended Haplotype Homozygosity

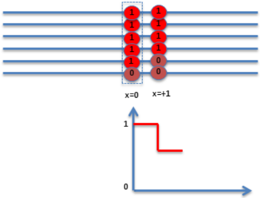


$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

$$EHH_c(x_i = +1) = ?$$

How many unique haplotypes carrying the core SNP?
What is their frequency?

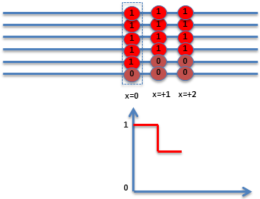
Extended Haplotype Homozygosity



$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

$$EHH_c(x_i = +1) = \frac{\binom{4}{2} + \binom{1}{2}}{\binom{5}{2}} = \frac{6+0}{10} = 0.60$$

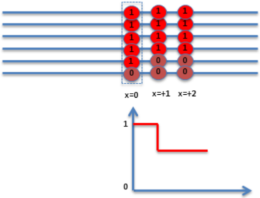
Extended Haplotype Homozygosity



$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

$$EHH_c(x_i = +2) = ?$$

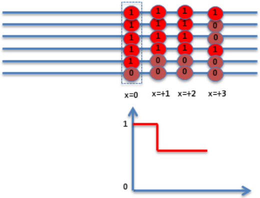
Extended Haplotype Homozygosity



$$EHH_c(x_i) = \sum_{h\in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

$$EHH_c(x_i = +2) = EHH_c(x_i = +1) = 0.60$$

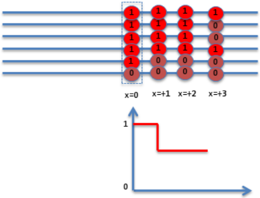
Extended Haplotype Homozygosity



$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

How many unique haplotypes carrying the core SNP?
What is their frequency?

Extended Haplotype Homozygosity

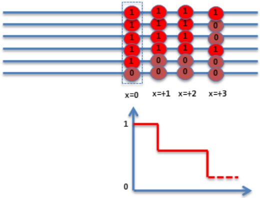


$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

How many unique haplotypes carrying the core SNP?
What is their frequency?
1111 with freq=2
1110 with freq=2
1000 with freq=1

$$EHH_c(x_i = +3) = ?$$

Extended Haplotype Homozygosity

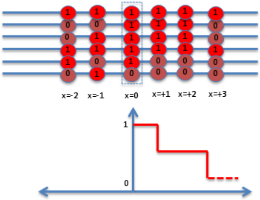


$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

How many unique haplotypes carrying the core SNP?
What is their frequency?
1111 with freq=2
1110 with freq=2
1000 with freq=1

$$EHH_c(x_i = +3) = \frac{\binom{2}{2} + \binom{2}{2} + \binom{1}{2}}{\binom{5}{2}} = \frac{1+1+0}{10} = 0.20$$

Extended Haplotype Homozygosity



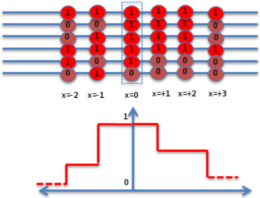
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

n	*n choose 2*
1	0
2	1
3	3
4	6
5	10
6	15

$$EHH_c(x_i = -1) = ?$$
$$EHH_c(x_i = -2) = ?$$

Comment on differences (if any) between *EHH(x=+2)* and *EHH(x=-2)*.

Extended Haplotype Homozygosity



$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

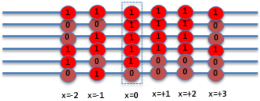
n	*n choose 2*
1	0
2	1
3	3
4	6
5	10
6	15

$$EHH_c(x_i = -1) = \frac{\binom{3}{2} + \binom{2}{2}}{\binom{5}{2}} = \frac{3+1}{10} = 0.4$$

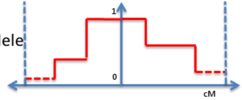
$$EHH_c(x_i = -2) = \frac{\binom{2}{2} + \binom{1}{2} + \binom{1}{2}}{\binom{5}{2}} = \frac{1+0+0}{10} = 0.1$$

Comment on differences (if any) between *EHH(x=+2)* and *EHH(x=-2)*?

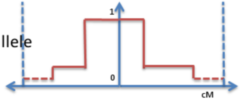
Integrated Haplotype Score



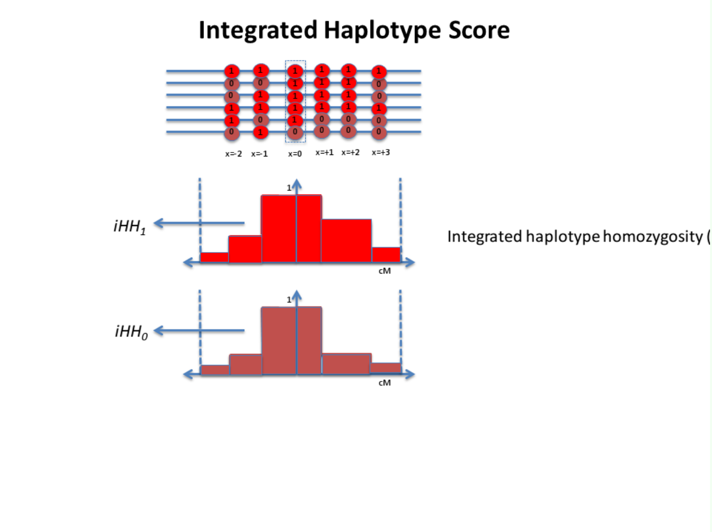
For the derived allele



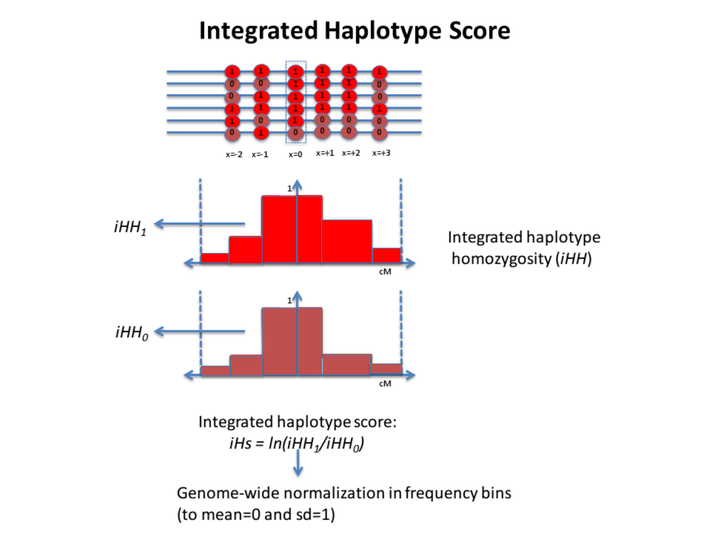
For the ancestral allele



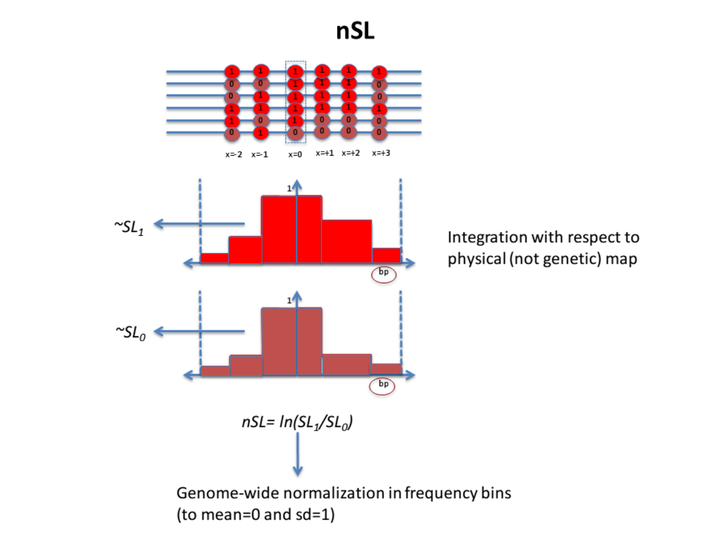
iHs


Integrated Haplotype Score
x=-2
x=-1
x=0
x=+1
x=+2
x=+3
iHH1
iHH0
cM
Integrated haplotype homozygosity (

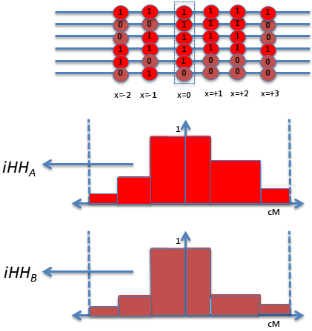
iHs



often $|iHs|$ is used


nSL
x=-2 x=-1 x=0 x=+1 x=+2 x=+3
$\sim SL_1$
1
bp
$\sim SL_0$
1
bp
Integration with respect to physical (not genetic) map
$nSL= ln(SL_1/SL_0)$
Genome-wide normalization in frequency bins (to mean=0 and sd=1)

Cross-population Extended Haplotype Homozygosity



Integrated haplotype homozygosity (*iHH*) for **populations** A and B

Integrated haplotype score:

$$XP\text{-}EHH = \ln(iHH_A/iHH_B)$$

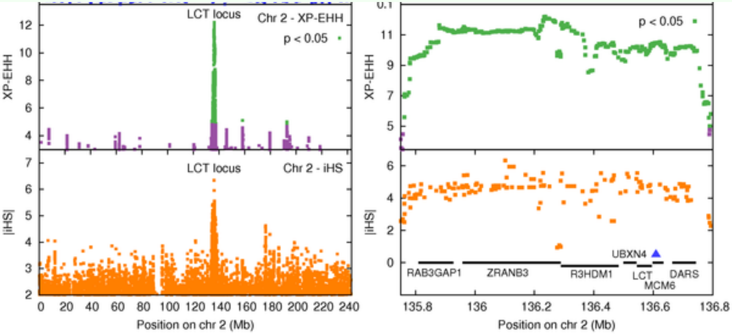
Genome-wide normalization in frequency bins (to mean=0 and sd=1)

Exercises

Let see how the haplotype methods works on famous examples of human adaptation (LCT).

go to

`http://popgen.dk/albrecht/BAG2018/web/`


LCT locus
Chr 2 - XP-EHH
p < 0.05
XP-EHH
Chr 2 - iHS
|iHS|
Position on chr 2 (Mb)
RAB3GAP1
ZRANB3
R3HDM1
UBXN4
LCT
MCM6
DARS