

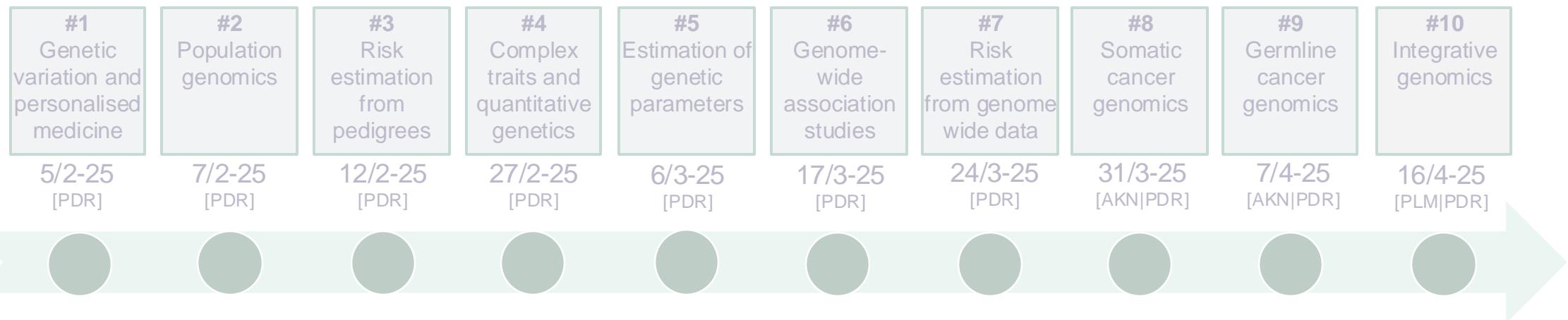
POPULATION GENOMICS

#2

PALLE DUUN ROHDE

palledr@hst.aau.dk

LETS GET STARTED



POPULATION GENOMICS

Today we will talk about

- Allele and genotype frequencies
- Hardy-Weinberg proportions
- Forces affecting genetic variation

OUTLINE

- 08:15 – 08:30** Recap
- 08:30 – 08:50** Lecture 1 [*Introduction to population genomics and frequencies*]
- 08:50 – 09:30** Break + Exercises Part 1 [E3, E4, E6]
- 09:30 – 09:50** Lecture 2 [*Hardy-Weinberg*]
- 09:50 – 10:30** Break + Exercises Part 2 [E8, E12]
- 10:30 – 10:50** Lecture 3 [*Modulation of genetic variation*]
- 10:50 – 11:45** Break + Exercises Part 3 [E13, E15]
- 11:45 – 12:00** Reflection



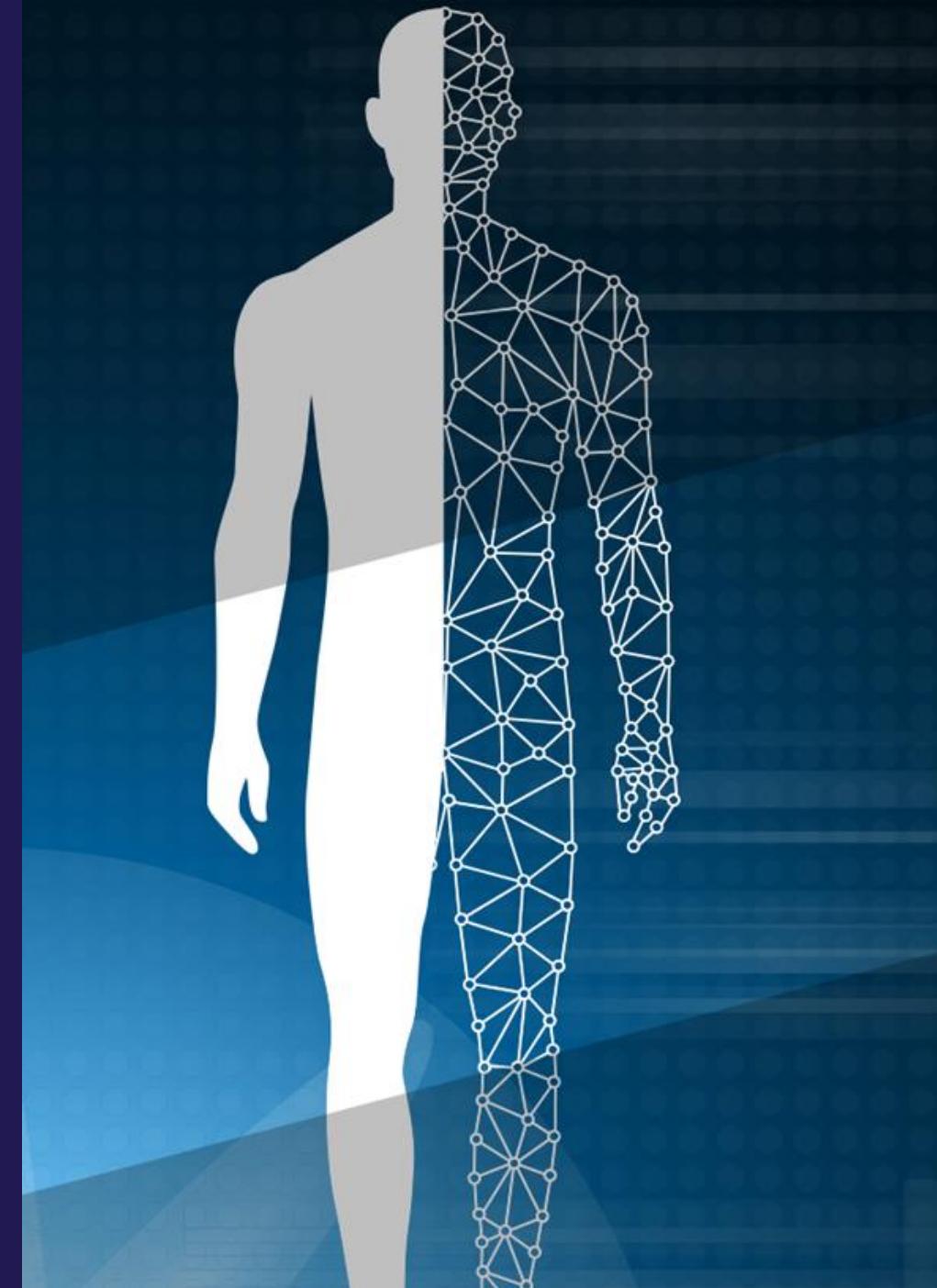
The remaining exercises
are also curriculum; thus,
you must do them on
your own.

OUTLINE

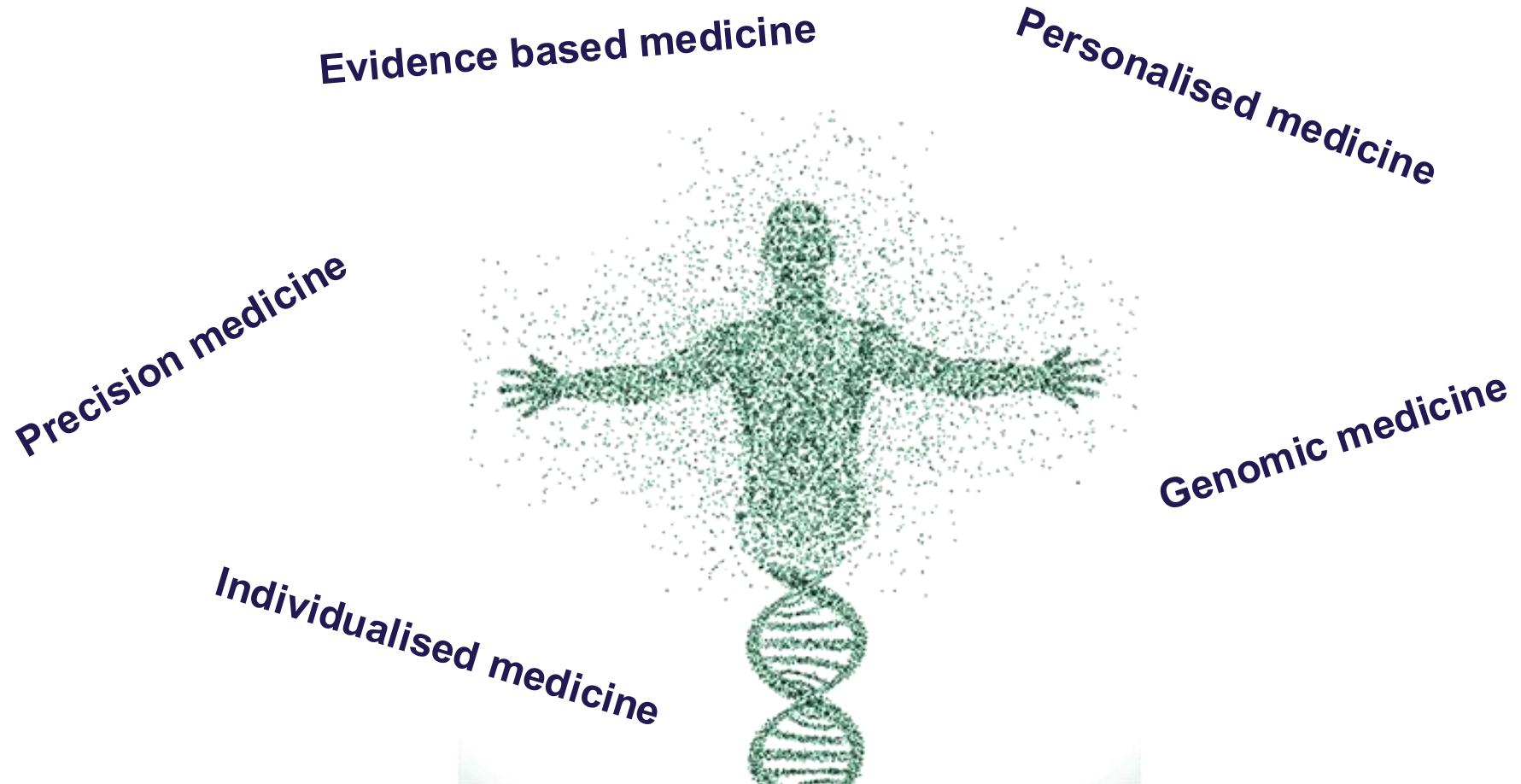
- | | |
|----------------------|--|
| 08:15 – 08:30 | Recap |
| 08:30 – 08:50 | Lecture 1 [<i>Introduction to population genomics and frequencies</i>] |
| 08:50 – 09:30 | Break + Exercises Part 1 [E3, E4, E6] |
| 09:30 – 09:50 | Lecture 2 [<i>Hardy-Weinberg</i>] |
| 09:50 – 10:30 | Break + Exercises Part 2 [E8, E12] |
| 10:30 – 10:50 | Lecture 3 [<i>Modulation of genetic variation</i>] |
| 10:50 – 11:45 | Break + Exercises Part 3 [E13, E15] |
| 11:45 – 12:00 | Reflection |

SHORT RECAP FROM LAST

- ❖ Personalised medicine
- ❖ Genetic variation



WHAT IS PERSONALISED MEDICINE?



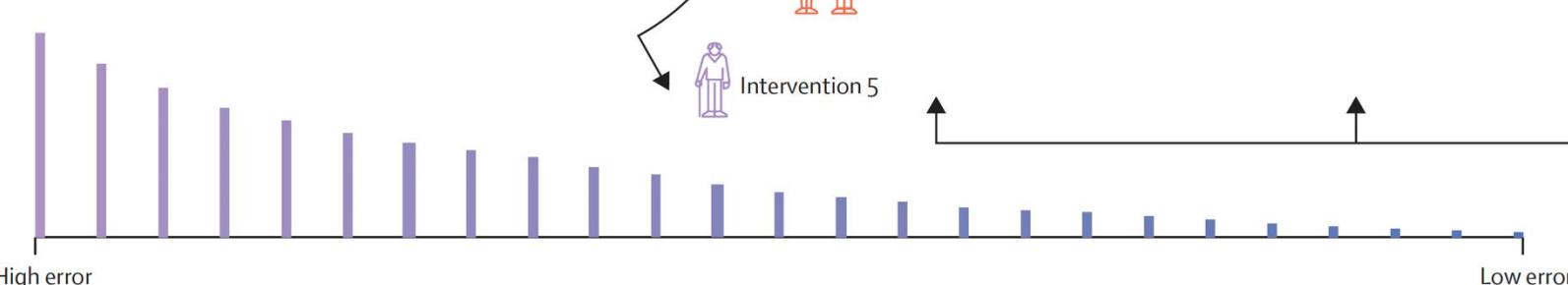
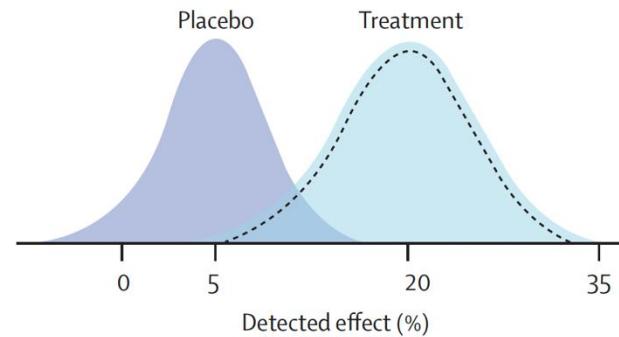
IMPLEMENTATION OF PRECISION MEDICINE

EPPOS [evidence-based precision personalised objective subjective]

Evidence-based Medicine

(1) Contemporary evidence-based medicine

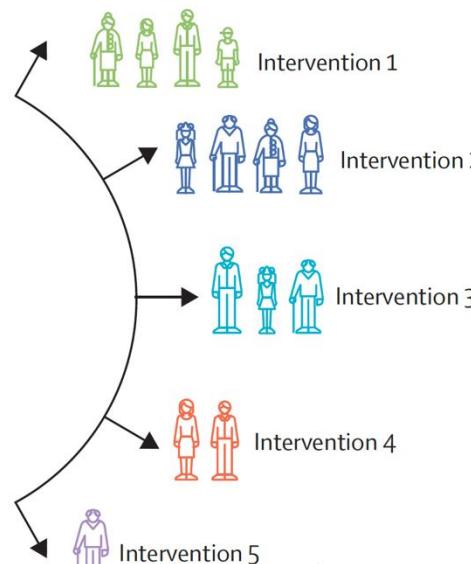
Estimate average risk or response using epidemiological and clinical trial cohorts



Precision Medicine

(2) Probability scoring and stratification

Maximise response and minimise risk using subclassification



Personalised Medicine

(3) Personalisation (objective)

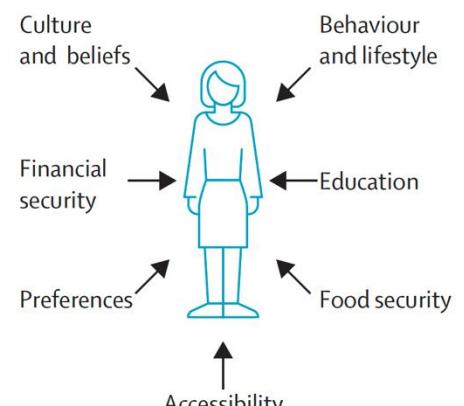
Monitor response to optimise dose, timing, and delivery



Individualised Medicine

(4) Personalisation (subjective)

Adapt intervention to fit the person's needs, capabilities, and preferences



WHY PERSONALISED MEDICINE?

Because people are different



- different disease risk
- respond differently to medication
- different side effects



Diagnostics, prognosis, treatment

FOCUS ON GENOMICS IN PRECISION MEDICINE

- 1) DNA is the *Blueprint* – identical from cradle-to-grave
- 2) Driven by *technological development*
- 3) One way causation [sickle cell disease]
- 4) A genetic test early in life have the potential to guide people
- 5) Other ‘omics also captures “environmental exposures”



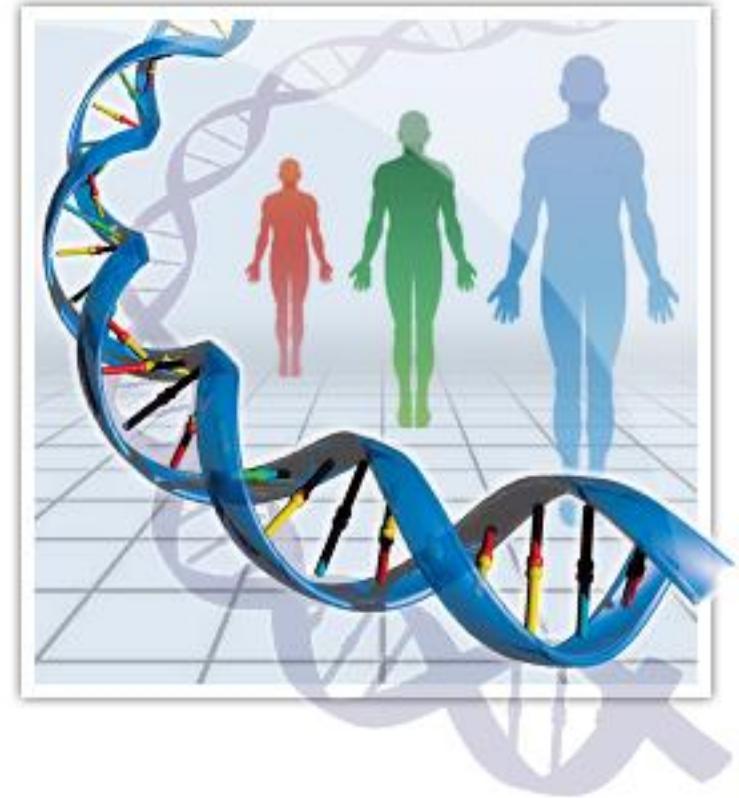
GENETIC DIVERSITY

Human evolution is driven by several different (evolutionary) factors

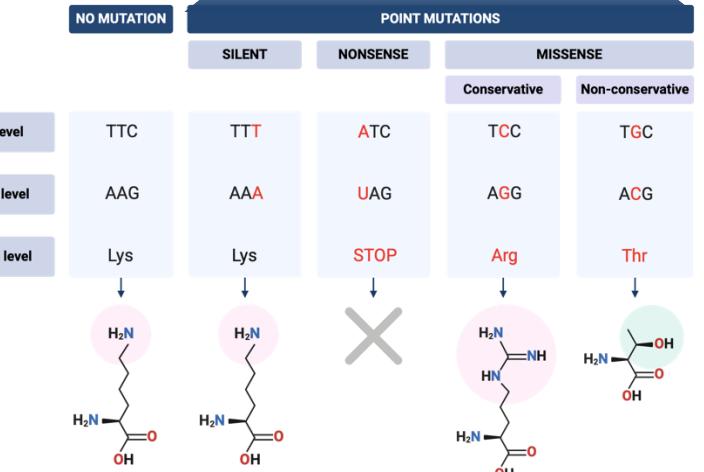
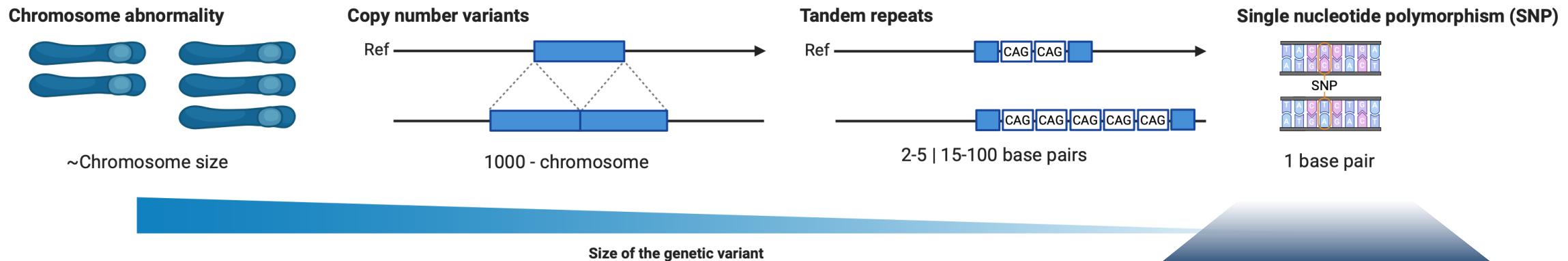
- ❖ Genetic mutations
- ❖ Migration
- ❖ Natural selection
- ❖ Genetic drift

The product is genetic diversity within a population.

Understanding the genetic diversity and how it has arisen is a necessary precursor to understand the genetics of complex traits.

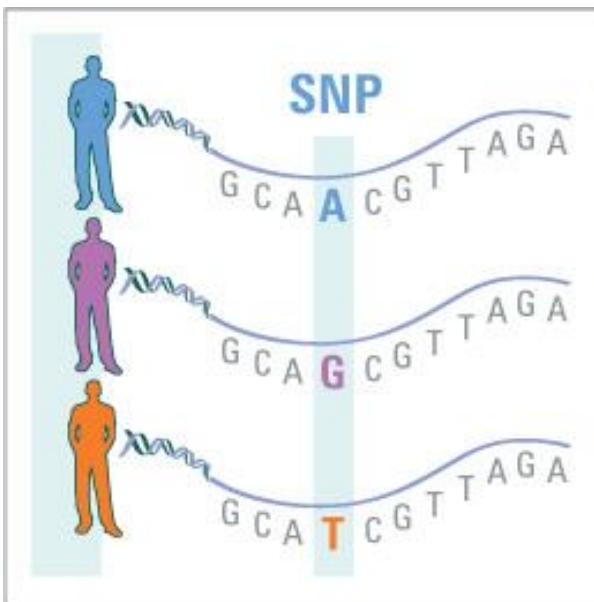


GENETIC VARIATION AT DIFFERENT RESOLUTION



GENETIC VARIATION

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)



Variation in the human genome

~3 billion base pairs

~90 million variants

Time

TaqMan

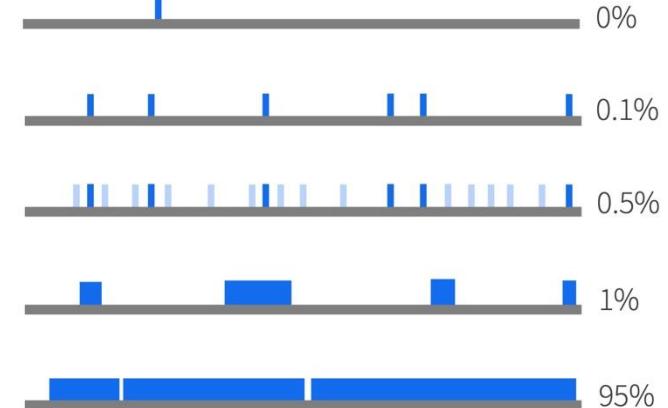
SNP array

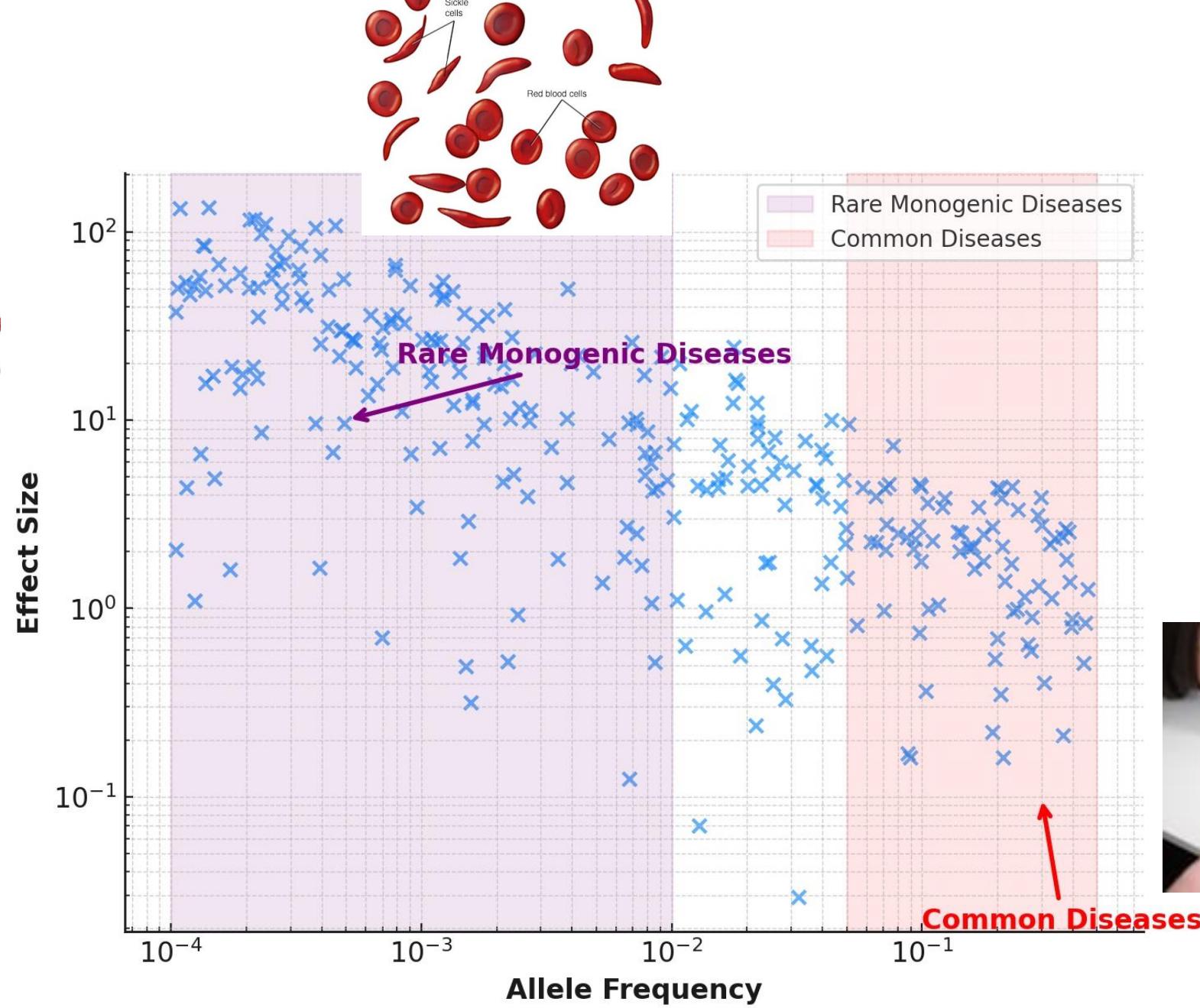
SNP array with imputation

Whole Exome Sequencing (WES)

Whole Genome Sequencing (WGS)

Genome coverage





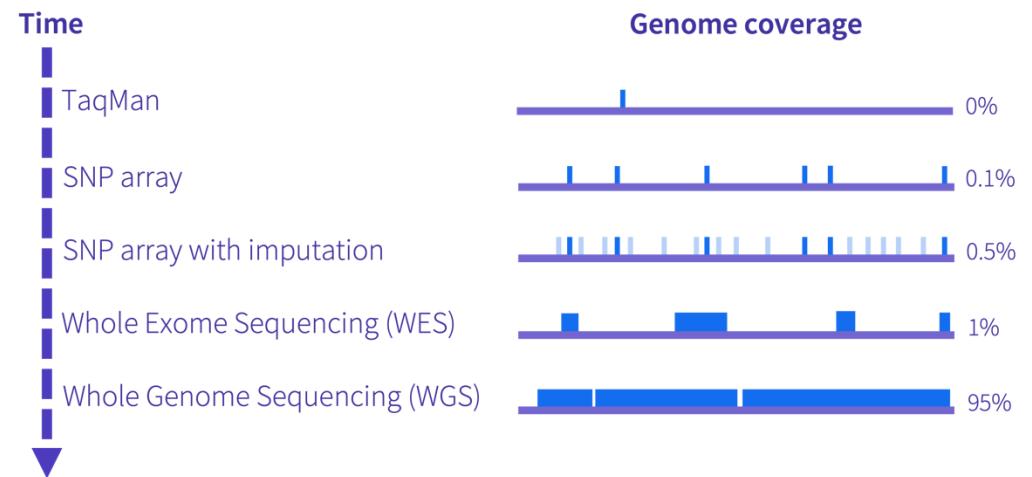
WHICH TECHNOLOGY?

The choice of technology to detect single nucleotide polymorphisms (SNPs) depends upon the application.

Clinical utility – WES might be preferred



For GWAS and PGS – genotyping is preferred



OUTLINE

- 08:15 – 08:30 Recap
- 08:30 – 08:50** Lecture 1 [*Introduction to population genomics and frequencies*]
- 08:50 – 09:30 Break + Exercises Part 1 [E3, E4, E6]
- 09:30 – 09:50 Lecture 2 [*Hardy-Weinberg*]
- 09:50 – 10:30 Break + Exercises Part 2 [E8, E12]
- 10:30 – 10:50 Lecture 3 [*Modulation of genetic variation*]
- 10:50 – 11:45 Break + Exercises Part 3 [E13, E15]
- 11:45 – 12:00 Reflection

POPULATION GENOMICS

The study of the distribution of hereditary variation across time and space in species and populations
[Bugge, F. 2008]



WHY IS POPULATION GENETICS IMPORTANT?

Population genomics tackles questions about genetic diversity

0.08% of nucleotide base pair in human DNA vary among individuals

Humans and orangutans are ~97% similar

Why this little genetic diversity?

- Selection favour functionally different DNA alleles in different circumstances
- DNA variation is tolerated when the alleles of a gene are functionally equivalent

The **aim of population genomics** is to model the dynamics of **evolutionary change within and between populations**.

THE FOUR FORCES

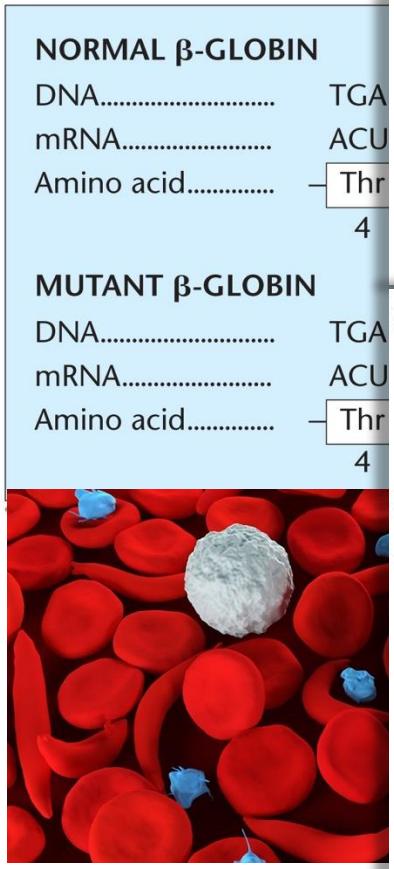
Mutation Copying errors during DNA replication, which introduce new alleles into the population

Natural selection differential transmission of alleles into the next generation due to the consequences of functional differences on an individual's survival and reproductive success

Genetic drift differential transmission of alleles into the next generation as a result of random sampling, and has the greatest potential impact in small populations

Gene flow spreads alleles from one population into another via migration, making them more genetically similar to each other, and countering genetic differentiation by drift

SIGNIFI FORCES



WHY IS POPULATION GENETICS IMPORTANT?

Population genetics tackles questions about genetic diversity

0.08% of nucleotide base pair in human DNA vary among individuals

Why this little genetic diversity?

- Selection favour functionally different DNA alleles in different circumstances
- DNA variation is tolerated when the alleles of a gene are functionally equivalent

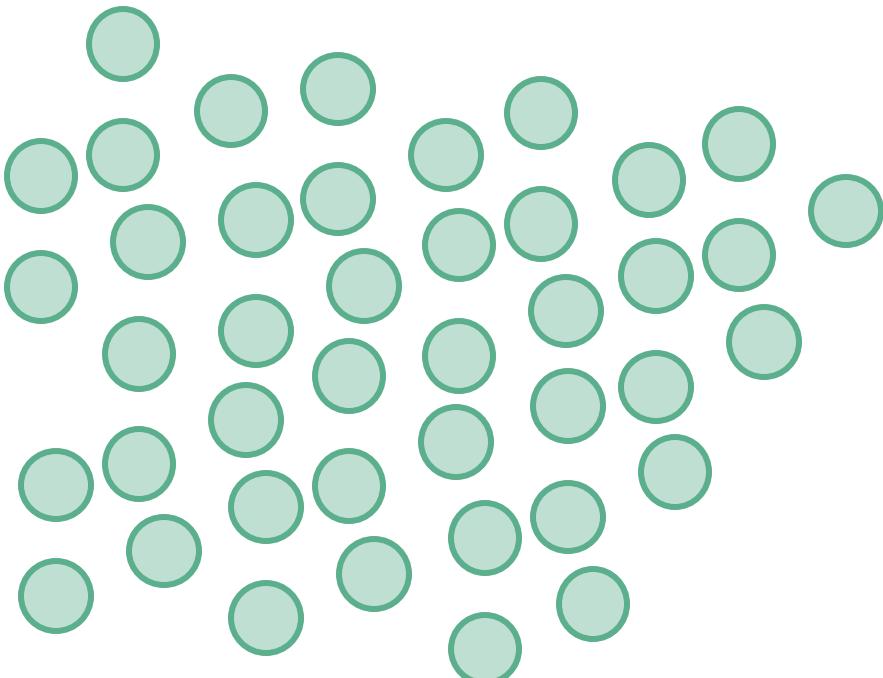
The **aim of population genetics** is to model the dynamics of **evolutionary change within and between populations**.

GENETIC VARIATION IN A SINGLE LOCUS



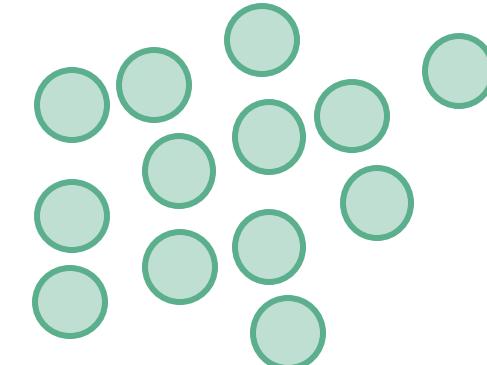
GENETIC VARIATION

IN A SINGLE LOCUS



A diploid ($2n$ alleles) population

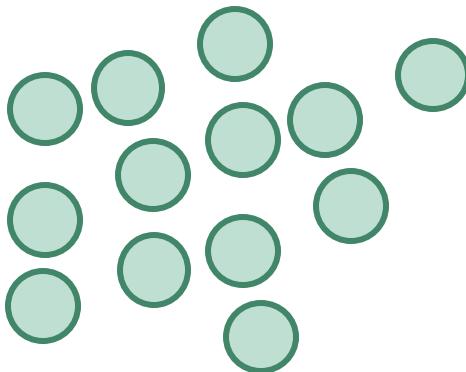
Random sampling



A random sample of individuals
of whom we know the genotype
of in a single locus

GENETIC VARIATION

IN A SINGLE LOCUS



A random sample of individuals
of whom we know the genotype
of in a single locus

Co-dominant (i.e., we can observe both alleles in heterozygote individuals).

The population is polymorph in one autosomal locus with the alleles **A** and **a**, and three genotypes, **AA**, **Aa** and **aa**.

The frequencies of the alleles are denoted **p** and **q**, and the frequency of the genotypes are **P_{AA}**, **P_{Aa}** and **P_{aa}**.

Note! There is a difference between \hat{p} and p . The hat ($\hat{}$) indicates that it is an estimate (\hat{p}) over the true parameter (p). For simplicity we ignore $\hat{}$.

FREQUENCIES

Genotype	AA	Aa	aa	Σ
Count	n_{AA}	n_{Aa}	n_{aa}	N
Genotype frequency	n_{AA}/N	n_{Aa}/N	n_{aa}/N	1

Allele frequency of A: $p = (2 \times n_{AA} + n_{Aa})/2 \times N$

We are counting the alleles

Allele frequency of a: $q = (2 \times n_{aa} + n_{Aa})/2 \times N$

Check! $p + q = 1$ ← All alleles are counted

EXAMPLE

MN blood group system is controlled by one locus with two co-dominant alleles L^M and L^N .

Genotype	MM	MN	NN	Σ
Count	64	120	16	200
Genotype frequency	$64/200$ $= 0.32$	$120/200$ $= 0.6$	$16/200$ $= 0.08$	1

$$\text{Allele frequency of } M: p = (2 \times n_{MM} + n_{MN}) / 2 \times N = \frac{(2 \times 64 + 120)}{(2 \times 200)} = 0.62$$

$$\text{Allele frequency of } N: q = (2 \times n_{NN} + n_{MN}) / 2 \times N = \frac{(2 \times 16 + 120)}{(2 \times 200)} = 0.38$$

Check $p + q = 0.62 + 0.38 = 1$

YOUR TURN



In a random sample of 100 individuals, we observe whether they can roll their tongue or not.

R = can roll tongue
r = cannot roll tongue

Genotype	RR	Rr	rr
Count	49	42	9

What is the frequency of the R allele?

YOUR TURN



In a random sample of 100 individuals, we observe whether they can roll their tongue or not.

R = can roll tongue
r = cannot roll tongue

Genotype	RR	Rr	rr
Count	49	42	9

THE ACCURACY OF FREQUENCIES

The accuracy of allele frequencies can be determined from their variances
- which are equal since $p = 1 - q$

Variance of p: $\text{Var}(p) = \frac{p(1-q)}{2N} + \frac{P_{AA}-p^2}{2N}$

Variance of p: $\text{Var}(p) = \frac{p(1-q)}{2N}$, if there are Hardy-Weinberg proportion (see later)

EXAMPLE

Genotype	AA	Aa	aa	Sum
Number	10	40	49	99
Frequency	0.101	0.404	0.495	1



Jaera ischiosetosa

$$\text{Allele frequency of A: } p = \frac{2 \times 10 + 40}{2 \times 99} = 0.303$$

$$\text{Allele frequency of a: } q = \frac{2 \times 49 + 40}{2 \times 99} = 0.697$$

$$\text{The variance of the allele frequency A: } \text{Var}(p) = \frac{0.303(1-0.303)}{2 \times 99} + \frac{0.101-0.303^2}{2 \times 99} = 0.00111$$

$$\text{The standard deviation of the allele frequency A: } \text{sd}(p) = \sqrt{0.00111} = 0.033$$

Assuming Gaussian distribution the 95% confidence interval is: estimate $\pm 1.96 \times \text{sd}$

$\rightarrow 0.303 [0.238-0.368]$

OUTLINE

08:15 – 08:30 Recap

08:30 – 08:50 Lecture 1 [*Introduction to population genomics and frequencies*]

08:50 – 09:30 Break + Exercises Part 1 [E3, E4, E6]

09:30 – 09:50 Lecture 2 [*Hardy-Weinberg*]

09:50 – 10:30 Break + Exercises Part 2 [E8, E12]

10:30 – 10:50 Lecture 3 [*Modulation of genetic variation*]

10:50 – 11:45 Break + Exercises Part 3 [E13, E15]

11:45 – 12:00 Reflection



OUTLINE

- 08:15 – 08:30 Recap
- 08:30 – 08:50 Lecture 1 [*Introduction to population genomics and frequencies*]
- 08:50 – 09:30 Break + Exercises Part 1 [E3, E4, E6]
- 09:30 – 09:50** **Lecture 2 [Hardy-Weinberg]**
- 09:50 – 10:30 Break + Exercises Part 2 [E8, E12]
- 10:30 – 10:50 Lecture 3 [*Modulation of genetic variation*]
- 10:50 – 11:45 Break + Exercises Part 3 [E13, E15]
- 11:45 – 12:00 Reflection

HARDY-WEINBERG LAW

So far, we have computed allele frequencies by counting genotypes

Genotype frequencies → Allele frequencies

Under certain conditions, we can compute genotype frequencies in the next generation

Allele frequencies → Genotype frequencies

However, that requires some assumptions.

THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation

Hardy-Weinberg principle describes the relationship allele- and genotype frequencies in the neutral population

HARDY-WEINBERG LAW

Known population parameters

AA	Aa	aa
P_{AA}	P_{Aa}	P_{aa}

What is the frequency in the next generation?

Using HW law

		Males		
		AA	Aa	aa
Females	AA	P_{AA}^2	$P_{AA}P_{Aa}$	$P_{AA}P_{aa}$
	Aa	$P_{Aa}P_{AA}$	P_{Aa}^2	$P_{Aa}P_{aa}$
	aa	$P_{aa}P_{AA}$	$P_{aa}P_{Aa}$	P_{aa}^2

HARDY-WEINBERG EQUILIBRIUM

		Genotypes of offspring		
Parental combinations	Frequency	AA	Aa	aa
AA x AA	P_{AA}^2	P_{AA}^2		
AA x Aa	$2 \times P_{AA}P_{Aa}$	$P_{AA}P_{Aa}$	$P_{AA}P_{Aa}$	
AA x aa	$2 \times P_{AA}P_{aa}$		$2 \times P_{AA}P_{aa}$	
Aa x Aa	P_{Aa}^2	$1/4 P_{Aa}^2$	$1/2 P_{Aa}^2$	$1/4 P_{Aa}^2$
Aa x aa	$2 \times P_{Aa}P_{aa}$		$P_{Aa}P_{aa}$	$P_{Aa}P_{aa}$
Aa x aa	P_{aa}^2			P_{aa}^2

	A	a
A	AA	Aa
A	AA	Aa

	A	a
A	AA	Aa
a	Aa	aa

HARDY-WEINBERG EQUILIBRIUM

		Genotypes of offspring		
Parental combinations	Frequency	AA	Aa	aa
AA x AA	P_{AA}^2	P_{AA}^2		
AA x Aa	$2 \times P_{AA}P_{Aa}$	$P_{AA}P_{Aa}$	$P_{AA}P_{Aa}$	
AA x aa	$2 \times P_{AA}P_{aa}$		$2 \times P_{AA}P_{aa}$	
Aa x Aa	P_{Aa}^2	$1/4 P_{Aa}^2$	$1/2 P_{Aa}^2$	$1/4 P_{Aa}^2$
Aa x aa	$2 \times P_{Aa}P_{aa}$		$P_{Aa}P_{aa}$	$P_{Aa}P_{aa}$
Aa x aa	P_{aa}^2			P_{aa}^2

$\sum AA = P_{AA}^2 + P_{AA}P_{Aa} + 1/4 P_{Aa}^2 = p^2$
 $= (P_{AA} + 1/2P_{Aa})^2 = p^2$

$\sum aa = P_{aa}^2 + P_{aa}P_{Aa} + 1/4 P_{Aa}^2 = q^2$
 $= (P_{aa} + 1/2P_{Aa})^2 = q^2$

$\sum Aa = P_{AA}P_{Aa} + 2P_{AA}P_{aa} + 1/2 P_{Aa}^2 + P_{Aa}P_{aa} = 2(P_{AA} + 1/2P_{Aa})(P_{aa} + 1/2P_{Aa}) = 2pq$

$$p = \frac{2N_{AA} + N_{Aa}}{2N} = \frac{N_{AA} + 1/2N_{Aa}}{N} = P_{AA} + 1/2P_{Aa}$$

$$q = \frac{2N_{aa} + N_{Aa}}{2N} = \frac{N_{aa} + 1/2N_{Aa}}{N} = P_{aa} + 1/2P_{Aa}$$

HARDY-WEINBERG EQUILIBRIUM

After one generation under HW assumptions the genotype frequencies will be in equilibrium:

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2

Allele frequencies do not change!

		Males	
		A (p)	a (q)
Females	A (p)	p^2	pq
	a (q)	pq	q^2

TESTING H-W PROPORTIONS

Genotype	AA	Aa	aa
Observed	N_{AA}	N_{Aa}	N_{aa}
Expected	$E_{AA} = p^2N$	$E_{Aa} = 2pqN$	$E_{aa} = q^2N$

$$\chi^2 = \frac{(N_{AA} - E_{AA})^2}{E_{AA}} + \frac{(N_{Aa} - E_{Aa})^2}{E_{Aa}} + \frac{(N_{aa} - E_{aa})^2}{E_{aa}}$$

General ≥ 2 alleles

$$\chi^2 = \sum_{i=1}^m \frac{(Obs - Exp)^2}{Exp}$$

P -value is obtained from χ^2 -distribution and degrees of freedom (df):

$$df = \frac{n(n-1)}{2}, n = \text{number of alleles}$$

EXAMPLE

HIV-1

HIV-1 is the virus giving AIDS.
 Being homozygote for the CCR5 mutation Δ32 protects against HIV-1 virus, whereas heterozygotes are susceptible, and the disease progress slowly.

Genotype	1/1	1/Δ32	Δ32/Δ32	Σ
Observed	79	20	1	100
Expected	p^2N	$2pqN$	q^2N	N

Allele frequency of 1: $p = \frac{2 \times 79 + 20}{2 \times 100} = 0.89$

Allele frequency of Δ32 : $q = \frac{2 \times 1 + 20}{2 \times 100} = 0.11$

Expected	$0.89^2 \times 100$ = 79.21	$2 \times 0.89 \times 0.11 \times 100$ = 19.58	$0.11^2 \times 100$ = 1.21	100
----------	--------------------------------	---	-------------------------------	-----

$$\chi^2 = \frac{(79 - 79.21)^2}{79.21} + \frac{(20 - 19.58)^2}{19.58} + \frac{(1 - 1.21)^2}{1.21} = 0.046$$



This population
is in HW
proportions

EXAMPLE

HIV-1

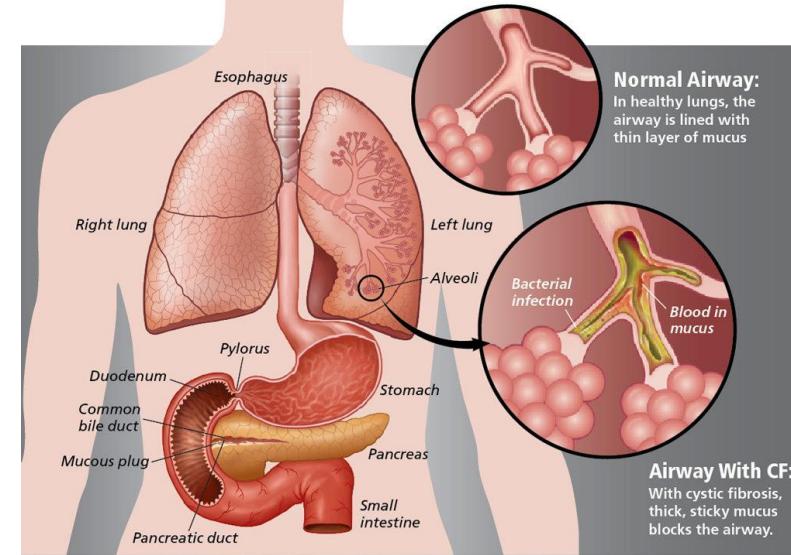


```
> NAA <- 79
> NAa <- 20
> Naa <- 1
> N <- NAA+NAa+Naa
>
> p <- (2*NAA+NAa)/(2*N)
> q <- (2*Naa+NAa)/(2*N)
>
> EAA <- p^2*N
> EAa <- 2*p*q*N
> Eaa <- q^2*N
>
> X <- (NAA-EAA)^2/EAA + (NAa-EAa)^2/EAa + (Naa-Eaa)^2/Eaa
> pchisq(q=X, df=1, lower.tail=FALSE)
[1] 0.8301536
>
```

YOUR TURN

Cystic fibrosis (CF) is a hereditary autosomal recessive disease, that, among other things, affects the lungs causing chronic/frequent lung infections.

In Europe, the prevalence of children born with cystic fibrosis (CF) is approximately 1/2500.



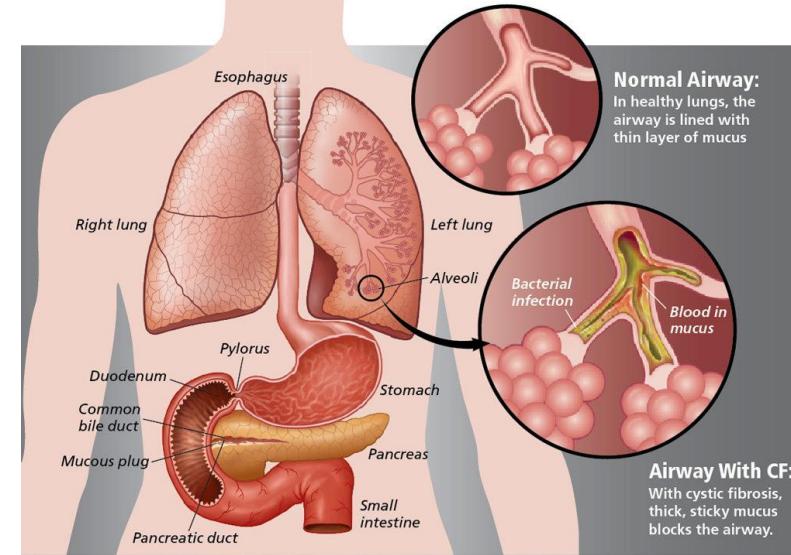
What is the frequency of the CF-allele?

Assume Hardy-Weinberg proportions.

YOUR TURN

Cystic fibrosis (CF) is a hereditary **autosomal recessive** disease, that, among other things, affects the lungs causing chronic/frequent lung infections.

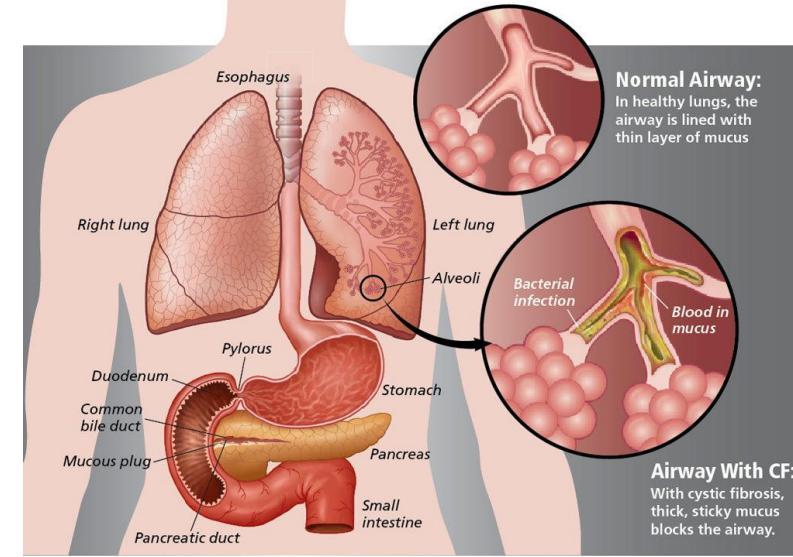
In Europe, the prevalence of children born with cystic fibrosis (CF) is approximately 1/2500.



YOUR TURN AGAIN

Cystic fibrosis (CF) is a hereditary autosomal recessive disease, that, among other things, affects the lungs causing chronic/frequent lung infections.

In Europe, the prevalence of children born with cystic fibrosis (CF) is approximately 1/2500.

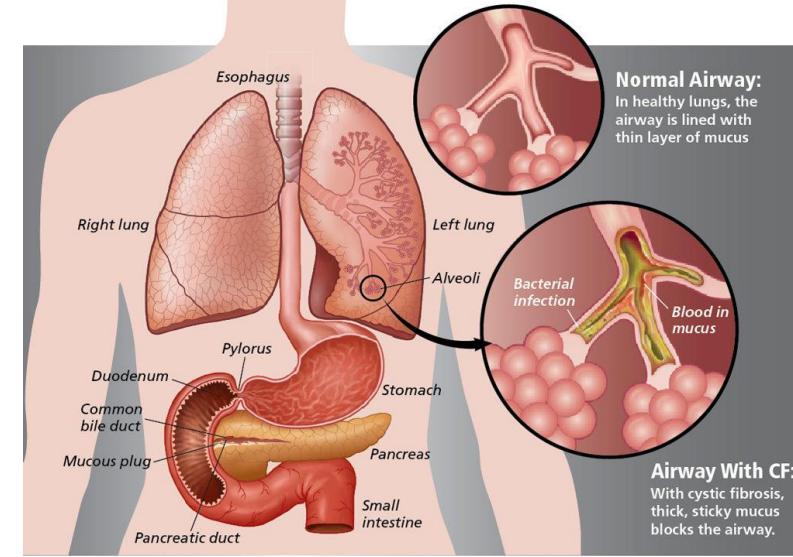


What is the frequency of healthy CF-carriers?

YOUR TURN AGAIN

Cystic fibrosis (CF) is a hereditary autosomal recessive disease, that, among other things, affects the lungs causing chronic/frequent lung infections.

In Europe, the prevalence of children born with cystic fibrosis (CF) is approximately 1/2500.



HARDY-WEINBERG PROPORTIONS

>2 ALLES

		Males		
		$A_1 (p_1)$	$A_2 (p_2)$	$A_3 (p_3)$
Females	$A_1 (p_1)$	$A_1A_1 (p_1^2)$	$A_1A_2 (p_1p_2)$	$A_1A_3 (p_1p_3)$
	$A_2 (p_2)$	$A_2A_1 (p_2p_1)$	$A_2A_2 (p_2^2)$	$A_2A_3 (p_2p_3)$
	$A_3 (p_3)$	$A_3A_1 (p_3p_1)$	$A_3A_2 (p_3p_2)$	$A_3A_3 (p_3^2)$

HARDY-WEINBERG PROPORTIONS

>2 ALLES

Genotype frequencies after random mating:

$$A_1A_1: p_1^2$$

$$A_2A_2: p_2^2$$

$$A_3A_3: p_3^2$$

$$A_1A_2: 2p_1p_2$$

$$A_1A_3: 2p_1p_3$$

$$A_2A_3: 2p_2p_3$$

Allele frequencies after random mating:

$$p_1 = p_1^2 + 0.5 \times 2p_1p_2 + 0.5 \times 2p_1p_3 = p_1(p_1 + p_2 + p_3) = p_1$$

$$p_2 = p_2^2 + 0.5 \times 2p_1p_2 + 0.5 \times 2p_2p_3 = p_2(p_1 + p_2 + p_3) = p_2$$

$$p_3 = p_3^2 + 0.5 \times 2p_1p_3 + 0.5 \times 2p_2p_3 = p_3(p_1 + p_2 + p_3) = p_3$$

$$p_1(p_1 + p_2 + p_3) = p_1(p_1 + (1 - p_1 - p_3) + p_3)$$

HARDY-WEINBERG PROPORTIONS

>2 ALLES

One locus with tre co-dominante alleles; A_1 , A_2 og A_3

Genotypes	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3	Σ
Observed	51	56	76	2	34	15	234

$$p_1 = \frac{2 \times 51 + 56 + 76}{2 \times 324} = 0.5$$

$$p_2 = \frac{2 \times 2 + 56 + 34}{2 \times 324} = 0.2$$

$$p_3 = \frac{2 \times 15 + 76 + 34}{2 \times 324} = 0.3$$

Expected	58.5	46.8	70.2	9.36	28.08	21.06	234
$\chi^2 = \sum \frac{(Obs-Exp)^2}{Exp}$	12.028						

$$\chi^2 = \sum \frac{(Obs-Exp)^2}{Exp} = 12.028, df = \frac{n(n-1)}{2} = \frac{3(3-1)}{2} = 3, p=0.0072$$



Deviation from HW proportions

VARIATION IN SEX-LINKED LOCI

When allele frequencies are **THE SAME** between males and females

		Males		
		X _A (p)	X _a (q)	Y (1)
Females	X _A (p)	X _A X _A (p^2)	X _A X _a (pq)	X _A Y (p)
	X _a (q)	X _a X _A (qp)	X _a X _a (q ²)	X _a Y (q)

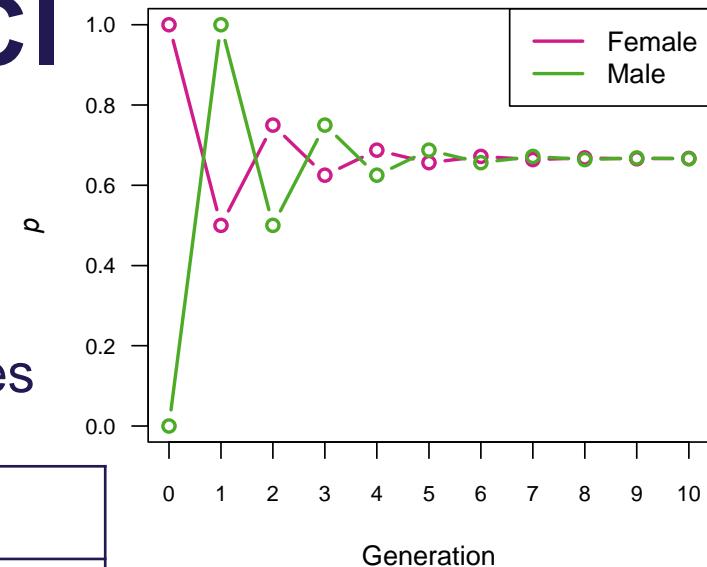
... then, the genotype frequency for males is the allele frequency.

VARIATION IN SEX-LINKED LOCI

When allele frequencies are **DIFFERENT** between males and females

		Males		
		$X_A (p^m)$	$X_a (q^m)$	$Y (1)$
Females	$X_A (p^f)$	$X_A X_A (p^f p^m)$	$X_A X_a (p^f q^m)$	$X_A Y (p^f)$
	$X_a (q^f)$	$X_a X_A (q^f p^m)$	$X_a X_a (q^f q^m)$	$X_a Y (q^f)$

At equilibrium: $p = (p^m + p^f)/3$



OUTLINE

- 08:15 – 08:30 Recap
- 08:30 – 08:50 Lecture 1 [*Introduction to population genomics and frequencies*]
- 08:50 – 09:30 Break + Exercises Part 1 [E3, E4, E6]
- 09:30 – 09:50 Lecture 2 [*Hardy-Weinberg*]
- 09:50 – 10:30 Break + Exercises Part 2 [E8, E12]
- 10:30 – 10:50 Lecture 3 [*Modulation of genetic variation*]
- 10:50 – 11:45 Break + Exercises Part 3 [E13, E15]
- 11:45 – 12:00 Reflection



OUTLINE

- 08:15 – 08:30 Recap
- 08:30 – 08:50 Lecture 1 [*Introduction to population genomics and frequencies*]
- 08:50 – 09:30 Break + Exercises Part 1 [E3, E4, E6]
- 09:30 – 09:50 Lecture 2 [*Hardy-Weinberg*]
- 09:50 – 10:30 Break + Exercises Part 2 [E8, E12]
- 10:30 – 10:50** **Lecture 3 [*Modulation of genetic variation*]**
- 10:50 – 11:45 Break + Exercises Part 3 [E13, E15]
- 11:45 – 12:00 Reflection

THE NEUTRAL POPULATION?

The **constancy of allele frequencies** from generation to generation only holds under the **assumptions of HW-law**.

- ⦿ Random mating
- ⦿ No selection
- ⦿ No genetic drift (infinite population size)
- ⦿ No migration
- ⦿ No mutation

Does the neutral population exists
?

THE NEUTRAL POPULATION

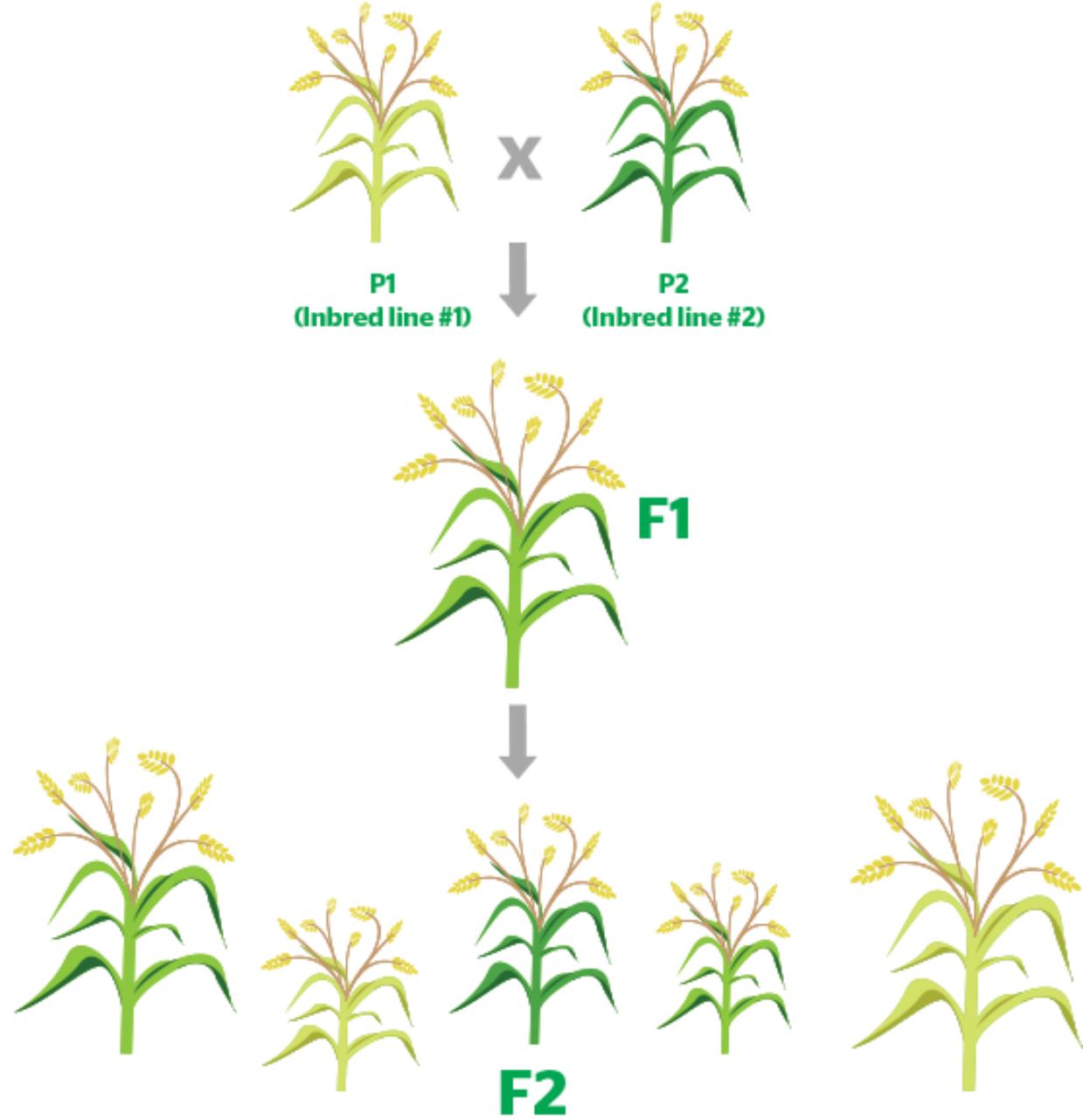
- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation

- Assortitative mating
- Isolation by distance
- Inbreeding



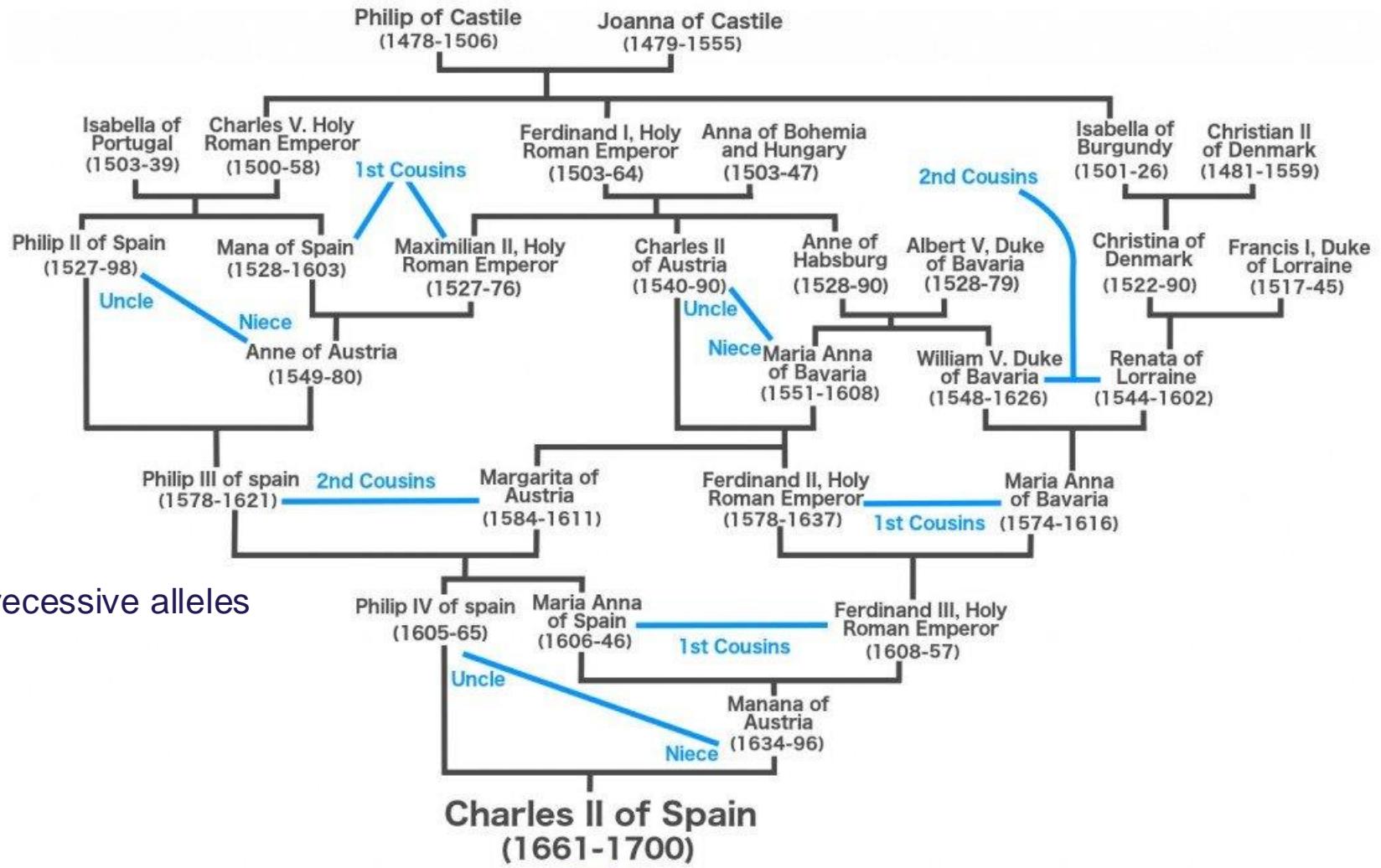
INBREEDING

- ⦿ Mating between relatives
 - ⦿ Heterosis | Hybrid vigor



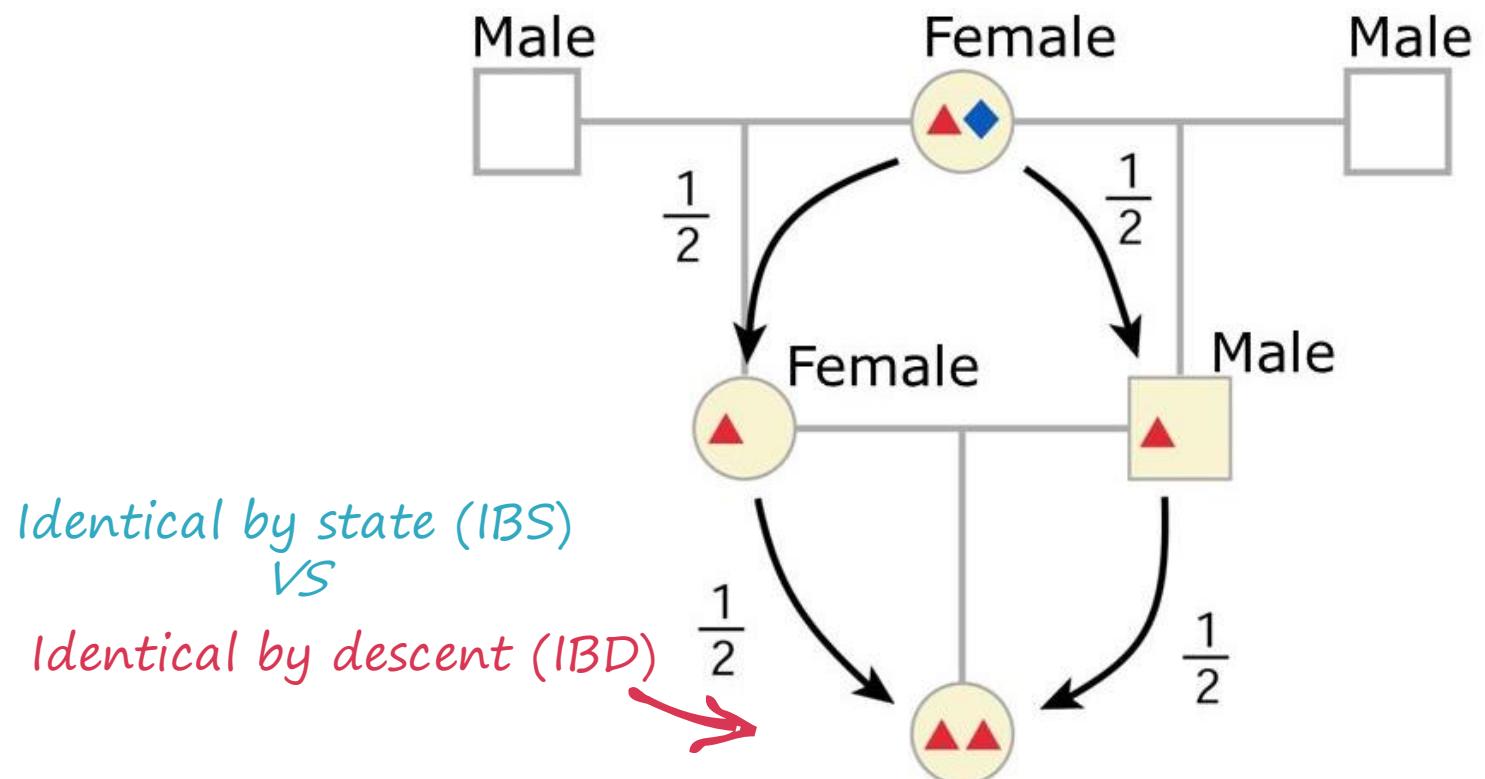
INBREEDING

- ⦿ Mating between relatives
- ⦿ Heterosis | Hybrid vigor
- ⦿ Inbreeding depression
 - › Accumulation of deleterious recessive alleles

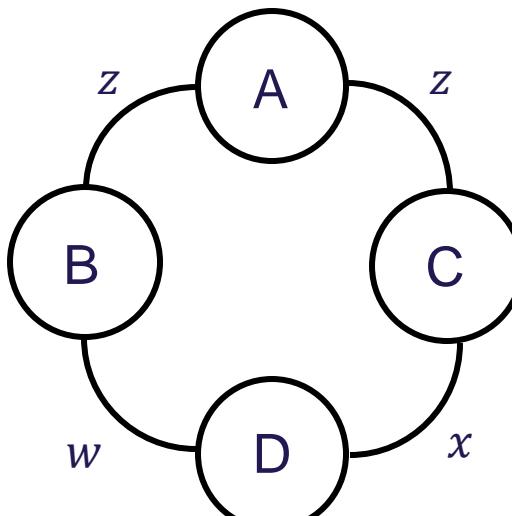
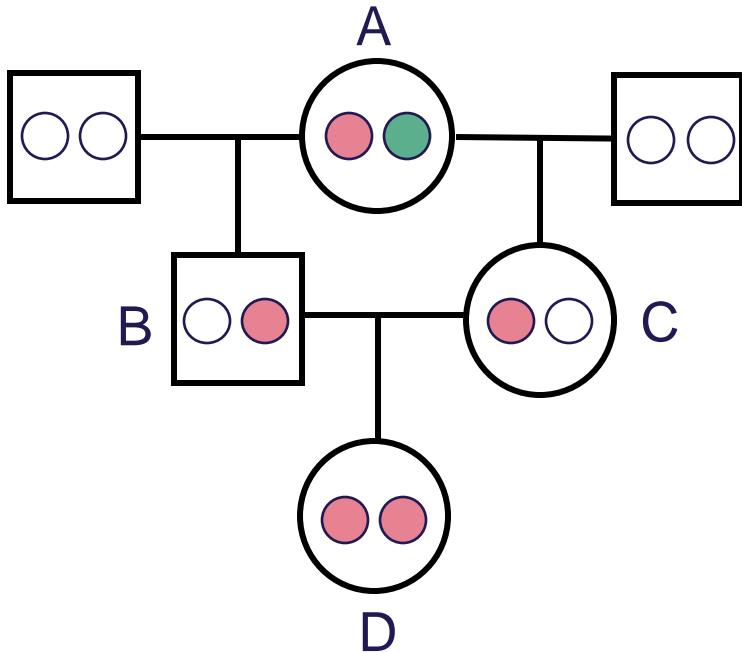


THE INBREEDING COEFFICIENT

The inbreeding coefficient (F) is the probability that two alleles in an individual trace back to the same copy in a common ancestor.



THE INBREEDING COEFFICIENT



Follow the transmission of alleles.

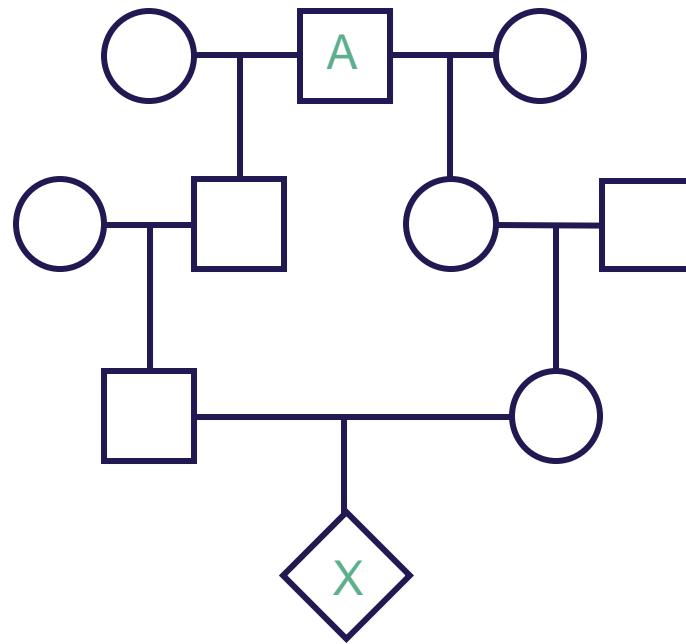
$$F_D = \left(\frac{1}{2}\right)^n (1 + F_A)$$

where *n* is the number of individuals in the loop without the individual we are computing *F* for.

$$F_D = \left(\frac{1}{2}\right)^3 (1 + F_A)$$

YOUR TURN

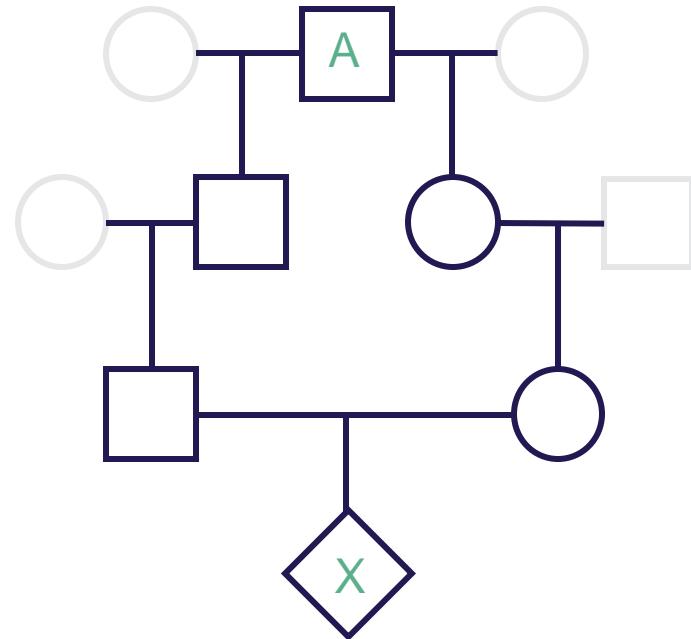
What is the inbreeding coefficient for individual **X** assuming individual **A** is not inbred ($F_A = 0$)?



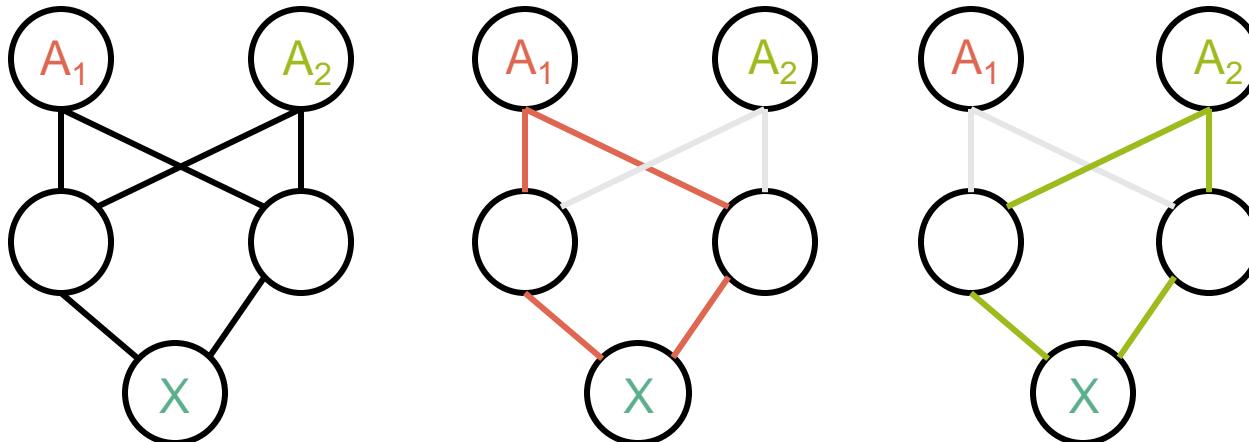
YOUR TURN

What is the inbreeding coefficient for individual **X** assuming individual **A** is not inbred ($F_A = 0$)?

$$F_X = \left(\frac{1}{2}\right)^5 (1 + 0) = \left(\frac{1}{2}\right)^5 = 0.031$$



WHEN THERE ARE MULTIPLE ANCESTORS



Follow the transmission of alleles over multiple loops.

$$F_X = \sum_{loops} \left(\frac{1}{2}\right)^n (1 + F_A)$$

INBREEDING

CHANGES GENOTYPE FREQUENCIES

If the population is in HW proportions

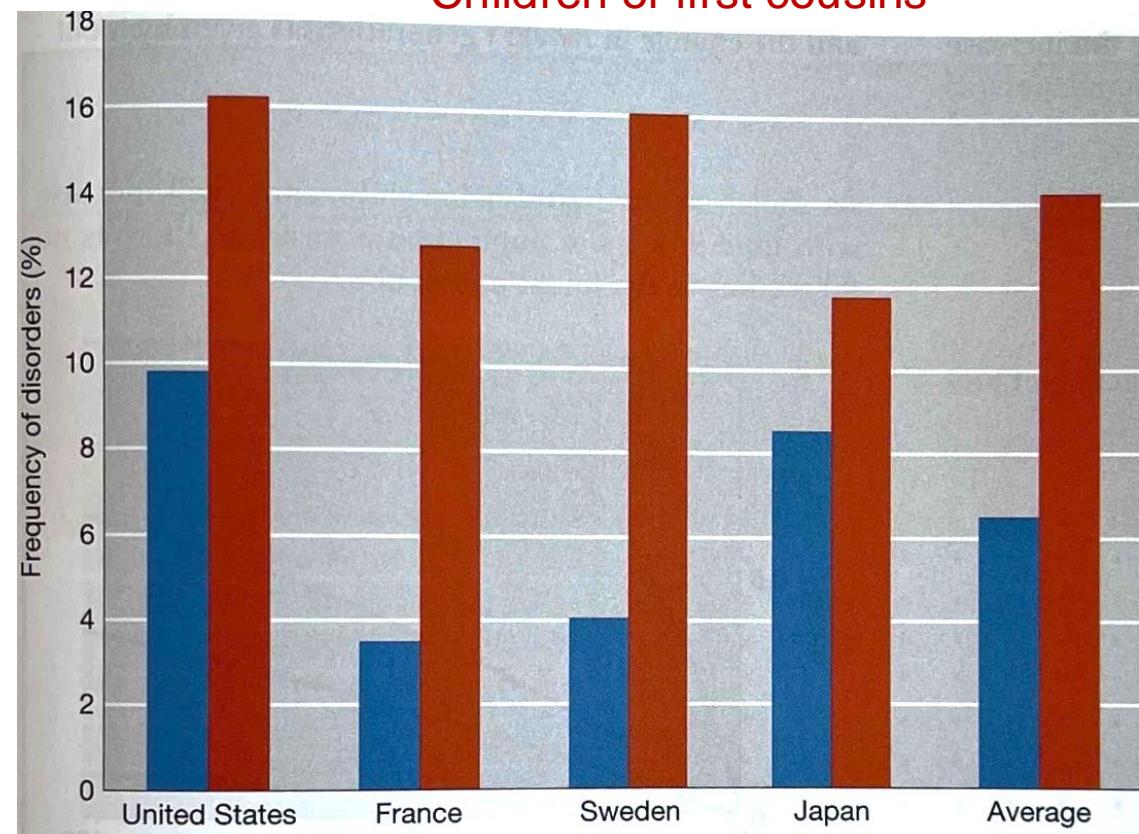
Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2

If there is inbreeding

Genotype	AA	Aa	aa
Frequency	p^2+pqF	$2pq-2pqF$	q^2+pqF

Results in excess in homozygotes

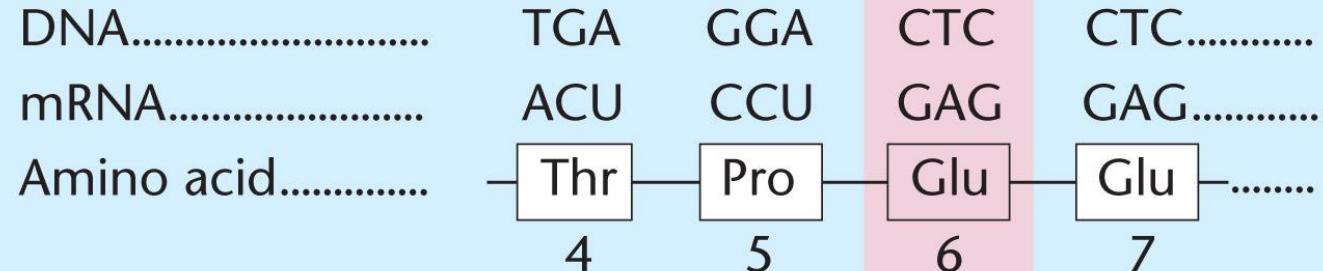
Children of unrelated parents
Children of first cousins



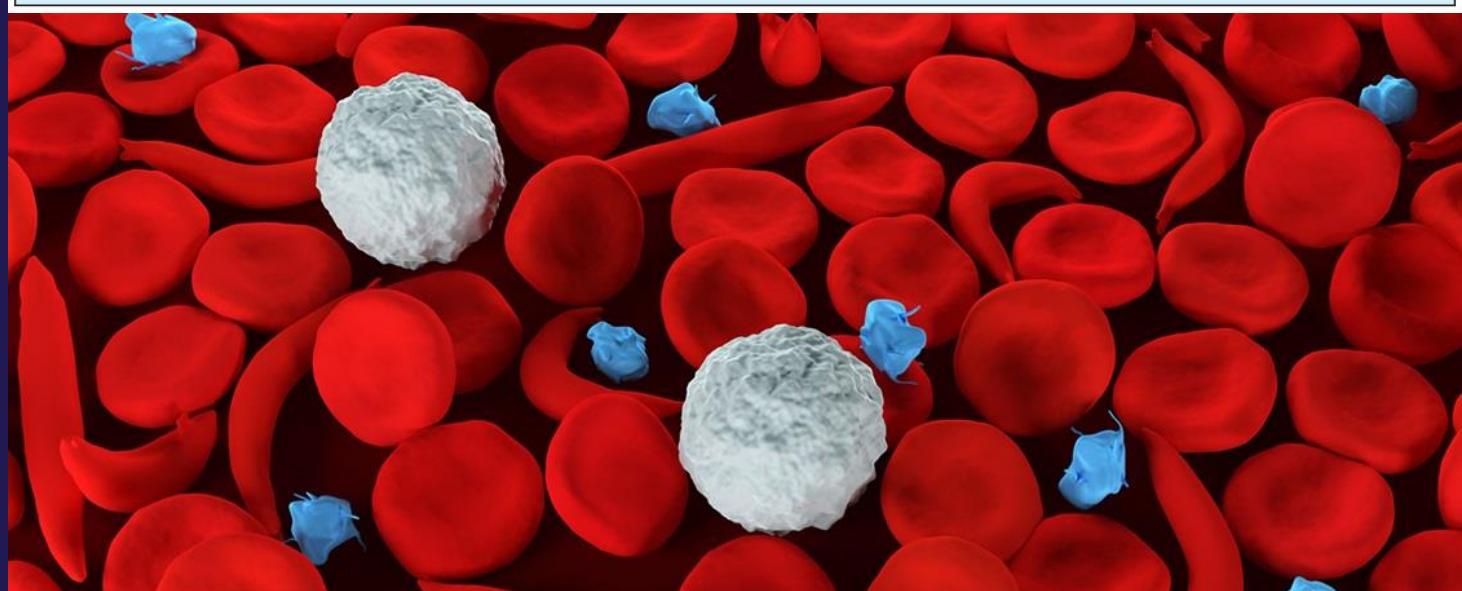
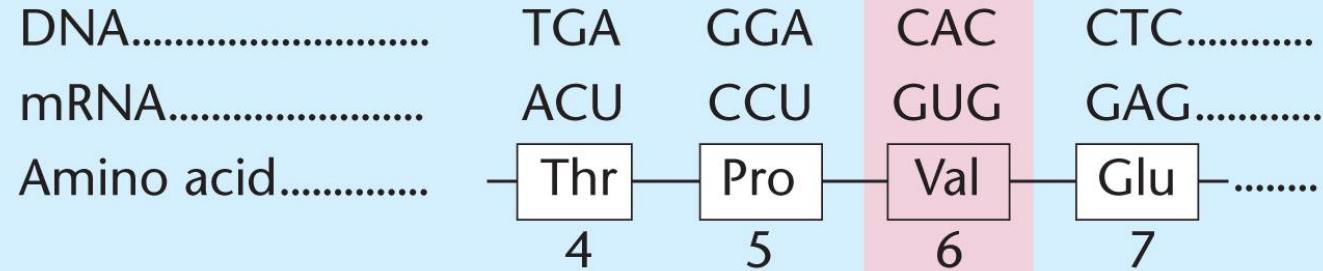
THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation**

NORMAL β -GLOBIN

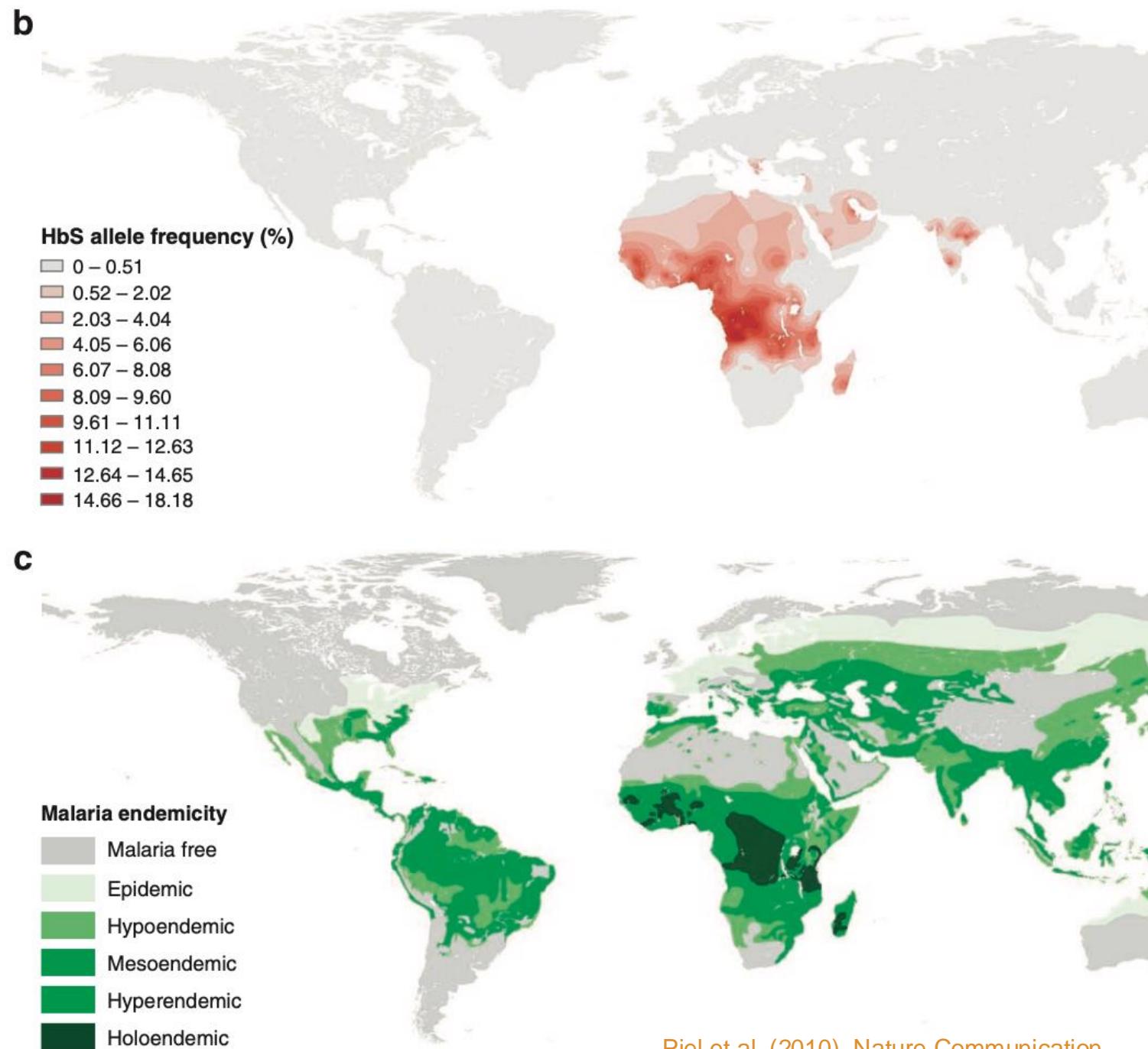
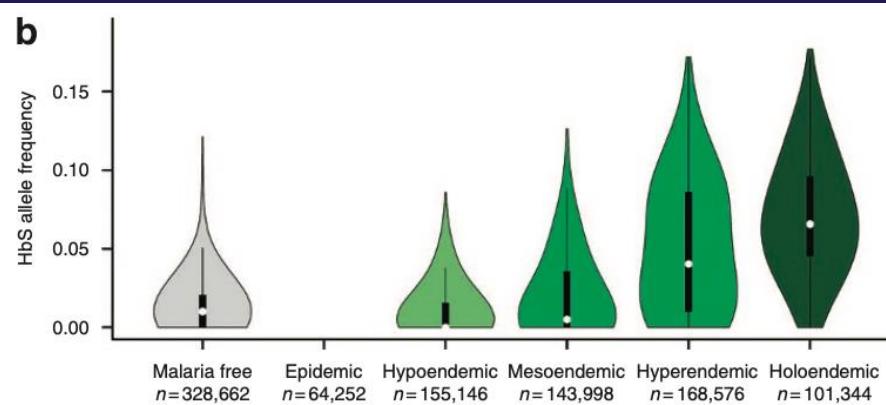


MUTANT β -GLOBIN



THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation



MUTATION AND SELECTION $a^+ \xrightarrow{\mu} a$

Number wildtype alleles in a population of $2N$ is $2Np$, which with the rate μ mutates to harmful allele.

In the next generation the proportion of new harmful alleles are: $\Delta q_\mu = 2Np\mu$

Recessive harmful

$$\Delta q_\mu = 2Nsq^2$$

$$q = \sqrt{\frac{\mu}{s}}$$

Genotype	a^+a^+	a^+a	aa
Fitness	1	1	$1-s$

Dominant harmful

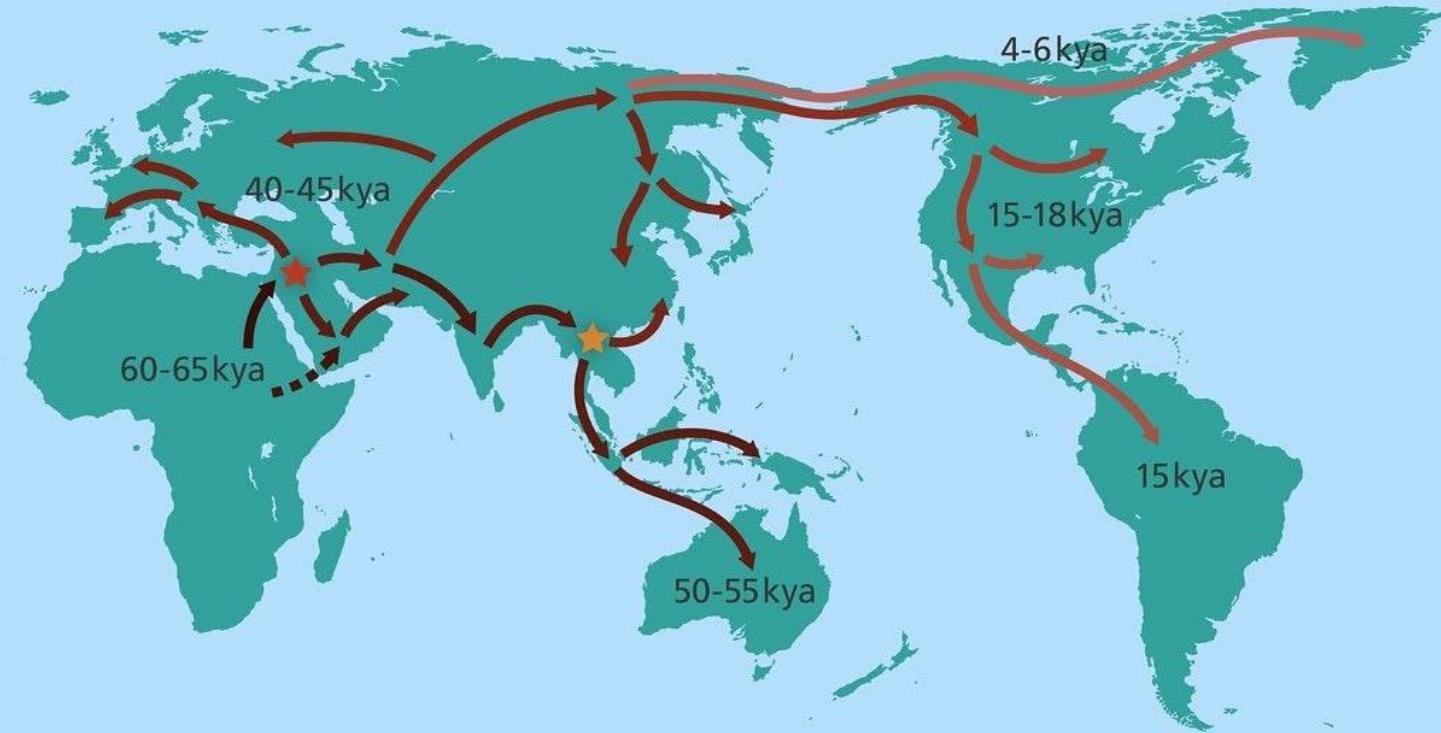
$$\Delta q_\mu = Ns2pq + 2Nsq^2$$

$$q = \frac{\mu}{s}$$

Genotype	a^+a^+	a^+a	aa
Fitness	1	$1-s$	$1-s$

THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift (infinite population size)
- **No migration**
- No mutation



···· alternative route

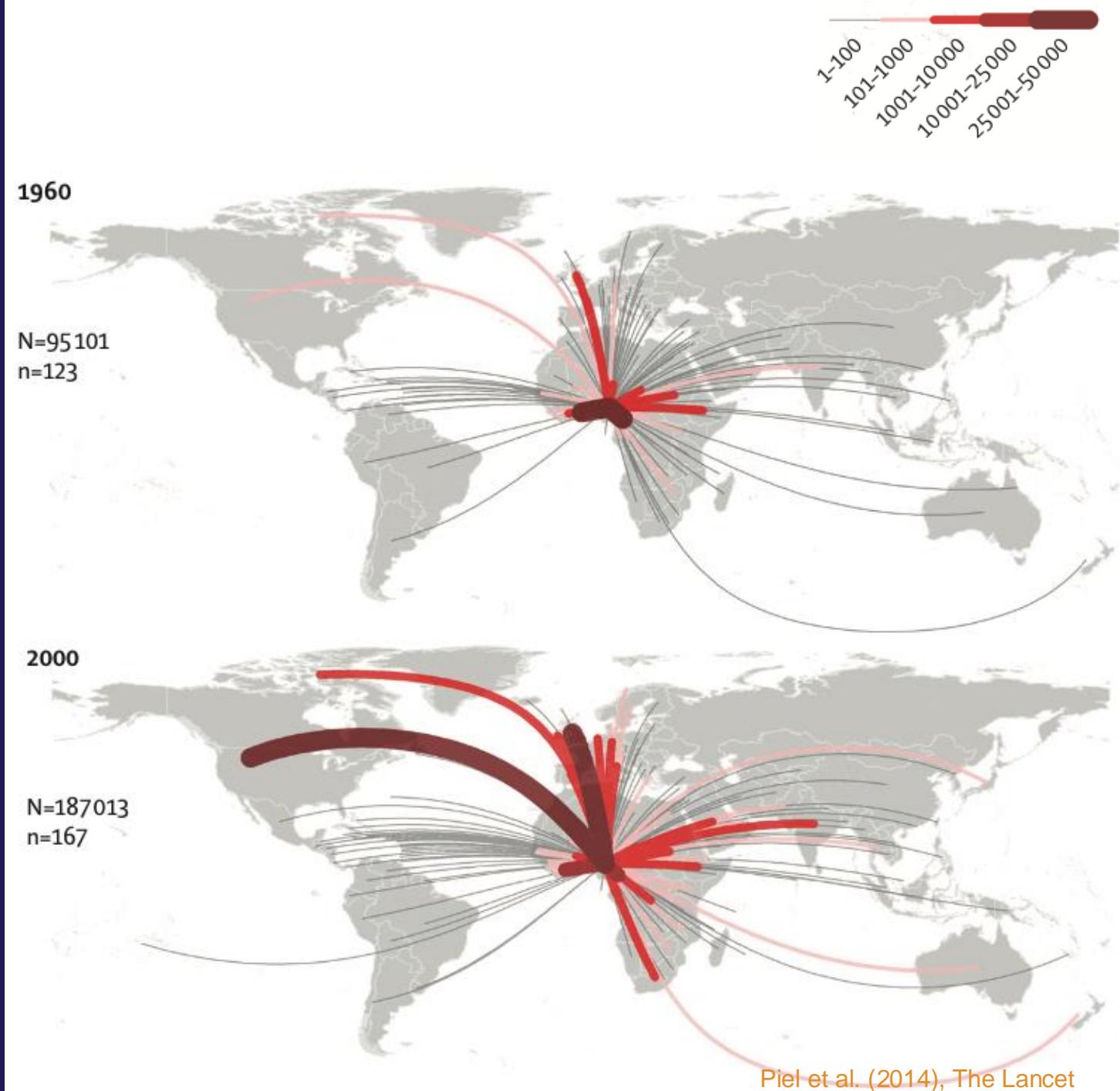
kya 1,000 years ago

★ possible location of admixture with Neanderthals

★ possible location of admixture with Denisovans

THE NEUTRAL POPULATION

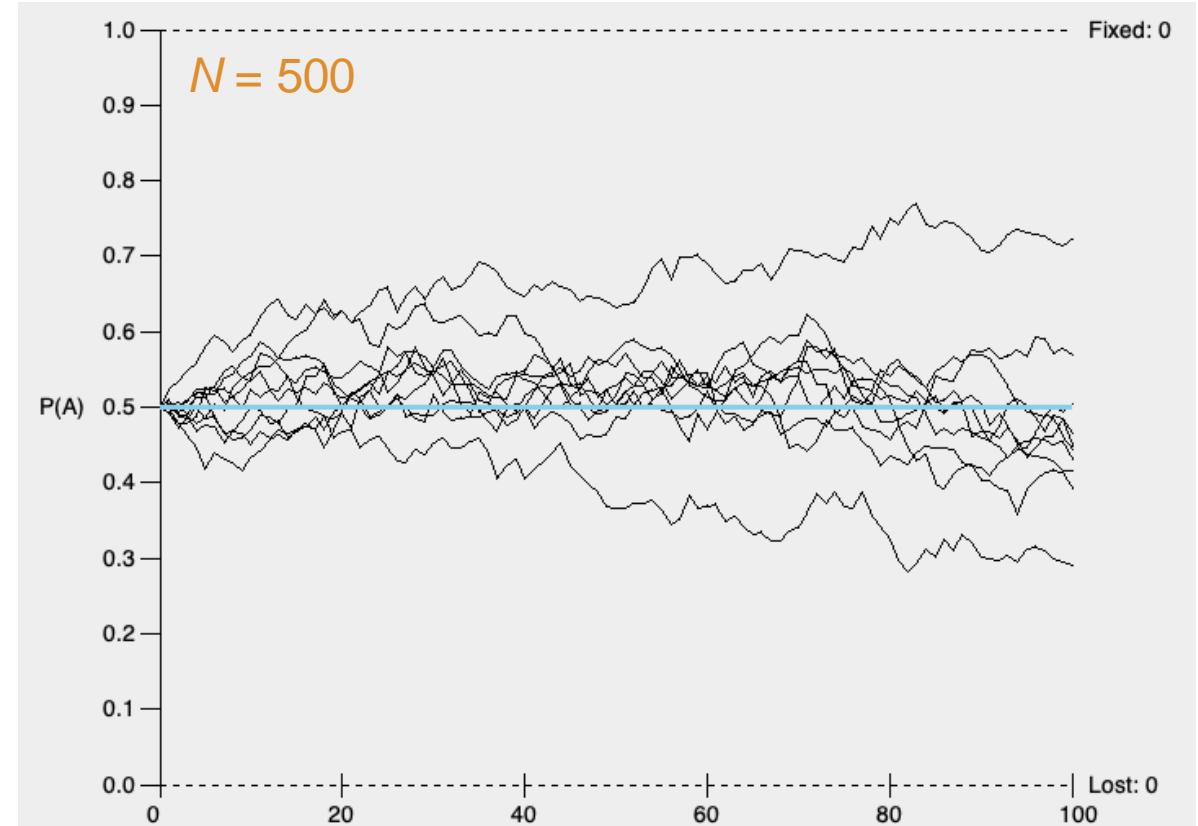
- Random mating
- No selection
- No genetic drift (infinite population size)
- **No migration**
- No mutation



THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift** (infinite population size)
- No migration
- No mutation

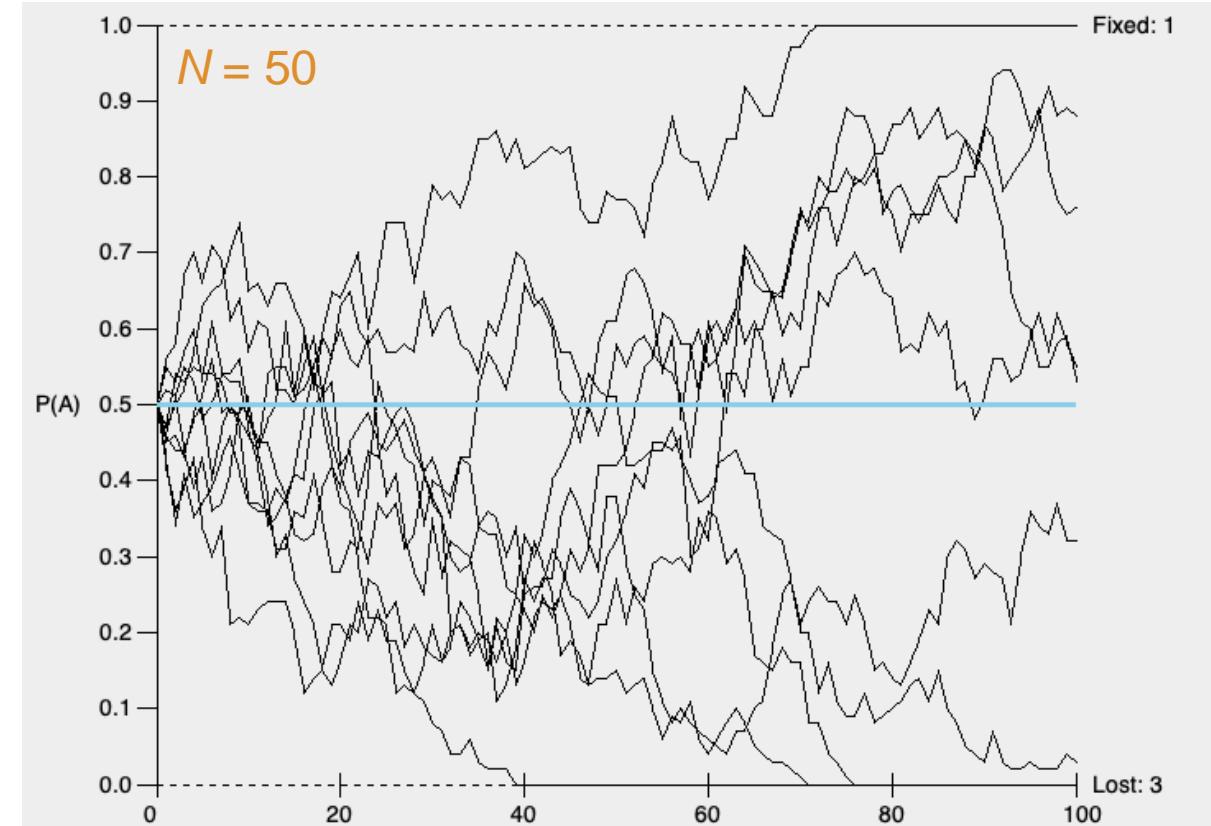
Genetic drift is **changes in allele frequencies between generations due to sampling error**



THE NEUTRAL POPULATION

- Random mating
- No selection
- **No genetic drift** (infinite population size)
- No migration
- No mutation

Genetic drift is **changes in allele frequencies between generations due to sampling error**



GENETIC DRIFT AND INBREEDING

Genetic drift entails loci in a sub-population becomes fixed, thus, the degree of homozygosity increases (thus, F increase).

The probability of selecting two gametes carrying the same allele is $1/(2N)$.

The degree of inbreeding increase with time

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

The rate of loss of heterozygosity (H) per generation

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0, \text{ the rate depend on } N$$

If there is inbreeding

Genotype	AA	Aa	aa
Frequency	$p^2 + pqF$	$2pq - 2pqF$	$q^2 + pqF$

Results in excess in homozygotes

MODULATION OF FREQUENCIES

Mutation	introduces new alleles <i>diversity within populations</i>
Migration	introduces new alleles <i>diversity within populations</i> <i>diversity between populations</i>
Genetic drift	loss of alleles <i>diversity within populations</i> <i>diversity between populations</i>
Selection	removes harmful alleles <i>diversity within populations</i> <i>diversity between populations</i>
Non-random mating	do not change alleles, but change genotype frequencies

OUTLINE

08:15 – 08:30 Recap

08:30 – 08:50 Lecture 1 [*Introduction to population genomics and frequencies*]

08:50 – 09:30 Break + Exercises Part 1 [E3, E4, E6]

09:30 – 09:50 Lecture 2 [*Hardy-Weinberg*]

09:50 – 10:30 Break + Exercises Part 2 [E8, E12]

10:30 – 10:50 Lecture 3 [*Modulation of genetic variation*]

10:50 – 11:45 Break + Exercises Part 3 [E13, E15]

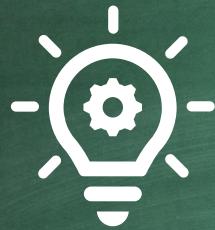
11:45 – 12:00 Reflection



OUTLINE

- 08:15 – 08:30** Recap
- 08:30 – 08:50** Lecture 1 [*Introduction to population genomics and frequencies*]
- 08:50 – 09:30** Break + Exercises Part 1 [E₃, E₄, E₆]
- 09:30 – 09:50** Lecture 2 [*Hardy-Weinberg*]
- 09:50 – 10:30** Break + Exercises Part 2 [E₈, E₁₂]
- 10:30 – 10:50** Lecture 3 [*Modulation of genetic variation*]
- 10:50 – 11:45** Break + Exercises Part 3 [E₁₃, E₁₅]
- 11:45 – 12:00** Reflection

REFLECT TOGETHER 2 AND 2



- What will you remember from today?
- What did you find difficult?
- What do you need to follow-up on?