HST.161 Molecular Biology and Genetics in Modern Medicine
Fall 2007

Harvard-MIT Division of Health Sciences and Technology
HST.161: Molecular Biology and Genetics in Modern Medicine, Fall 2007
Course Directors: Prof. Anne Giersch, Prof. David Housman

THALASSEMIAS

- PATHOLOGY IN THALASSEMIA IS A CONSEQUENCE OF AN IMBALANCE IN ALPHA AND BETA GLOBIN CHAIN SYNTHESIS

- EXCESS ALPHA OR BETA CHAINS ARE INSOLUBLE IN THE RBC

- PRECIPITATED GLOBIN GENES DAMAGE THE RED CELL MEMBRANE SHORTENING RED CELL HALF LIFE

- ANEMIA CAN BE CORRECTED BY TRANSFUSION

- CONTINUOUS TRANSFUSION CAN LEAD TO IRON OVERLOAD

Hemoglobinopathies and Thalassemias

- Mutations which alter the function of either the alpha or beta globin genes
- Hemoglobinopathies--mutations which cause a change in primary structure of one of the globin chains--over 700 known
- Thalassemias--mutations which alter the level of expression of one of the globin chains-- over 280 known

Mutations in β thalassemia

Image removed due to copyright restrictions.
Genetic map of mutations that cause beta-thalassemia.

β thalassemia alleles have been been described for almost every process affecting gene expression

α and β chromosomal loci

Image removed due to copyright restrictions.

Developmental expression pattern proceeds along
Each locus: ε→γ→β and ζ→α
Strong expression requires upstream regulatory elements

Images removed due to copyright restrictions.
Illustrations of alpha-globin gene cluster on chromosome 16, and beta-globin gene cluster on chromosome 11.

Thalassemia Genotypes and Syndromes

Alpha Thalassemia	α genes	Globin Chains	Hemoglobin	Anemia
Normal	$\alpha\alpha/\alpha\alpha$	$\alpha_2\beta_2$	A	None
Silent Carrier	$\alpha\alpha/\alpha\text{-}$	$\alpha_2\beta_2$	A	None
Trait	$\alpha\text{-}/\alpha\text{-}$ $\text{--}/\alpha\alpha$	$\alpha_2\beta_2$	A	Mild
Hb H disease	$\text{--}/\text{-}\alpha$	$\alpha_2\beta_2, \beta_4$	A, H	Intermediate
Hydrops fetalis	$\text{--}/\text{--}$	$\gamma_4,$ $\zeta_2\gamma_2$	Barts Portland	Lethal
Beta Thalassemia	β genes	Globin Chains	Hemoglobin	Anemia
Normal	β/β	$\alpha_2\beta_2$	A	None
Thalassemia minor	β^+/β β^o/β	$\alpha_2\beta_2, \alpha_2\delta_2, \alpha_2\gamma_2$	A, A_2, F	Mild
Thalassemia major	β^+/β^+ β^o/β^o	$\alpha_2\beta_2, \alpha_2\delta_2, \alpha_2\gamma_2,$ $\alpha_2\gamma_2, \alpha_2\delta_2$	A, A_2, F F, A_2	Severe Severe
HPFH*	γ/γ	$\alpha_2\gamma_2$	F	Mild

Figure by MIT OpenCourseWare.

Splice variants are one of the most frequent causes of β thalassemia alleles: A Mutation Can Create A New Splice Site in an Intron

Image removed due to copyright restrictions.

LOSS OF A NORMAL SPLICE SITE DUE TO A SINGLE BASE CHAGE

Common Meditarranean Mutation for β Thalassemia IVS-1, position 1 (G → A)

Image removed due to copyright restrictions.
Genetic sequence of common Mediterranean mutation for beta thalassemia.

MUTATION OF G TO A DESTROYS THE NORMAL SPLICE SIGNAL ADJACENT TO CODON 30; AN ABNORMAL mRNA IS PRODUCED WHICH INCLUDES SEQUENCES FROM INTRON 1; INCORRECT AMINO ACIDS ARE ADDED AFTER POSITION 30 AND A SHORT POLYPEPTIDE IS PRODUCED FOLLOWING A TERMINATION CODON WHICH OCCURS IN THE INTRON 1 SEQUENCE

Hemoglobin E: A Mutation in an Exon Creates a New Splice Site Causing 40% of mRNA made to be Non-functional;
 the remaining RNA encodes a β Globin with a single amino acid change--glutamic acid (GAG) to lysine (AAG) at position 26

Image removed due to copyright restrictions.

HbE has a very high allele frequency and widespread distribution in southeast Asia

Mutations in the Promoter, the 3' UTR or the poly A Site Can Reduce mRNA Expression Levels

Image removed due to copyright restrictions.

The TATA box is an important sequence for most eukaryotic promoters because it binds the key transcription factor TBP

Image removed due to copyright restrictions.

β Thalassemia mutations in the TATA box include:

-31 A to G

-30 T to A and -30 T to C

-29 A to G

-28 A to G

An important CCCCC element is located upstream of the start of transcription between positions -86 and -90 of the β globin gene

β Thalassemia mutations in this element include:
-90 C to T
-88 C to A or T
-87 C to A or G or T
-86 C to G

Mutations affecting mRNA polyadenylation at the polyA site can cause β thalassemia

- AATAAA is the β globin poly A site
- *Mutations seen in β thalassemia*
- AA*C*AAA
- AAT*T*AA
- AATT*G*A
- AATAA*C*

Image removed due to copyright restrictions.

Image removed due to copyright restrictions.

Capping the 5′End.
Caps at the 5′end of eukaryotic mRNA include 7-methylguanylate (red) attached by a triphosphate linkage to the ribose at the 5′end.

Mutation in silent β thalssemia +1 A to C prevents cap formation

A Mutation in the Chain termination Codon Causes a Longer α Globin to be Produced; the Mutation also Causes Instability of the mRNA Leading to Reduced levels of Gene Expression

Image removed due to copyright restrictions.
Chart showing that reduced mRNA levels correlate with reduced globin expression.

DNA sequences such as β LCR and α HS40 play a Key Role in Controlling Expression of Each Locus

Image removed due to copyright restrictions.

DELETION OF THE HS40 BOX LEADS TO INACTIVATION OF TRANSCRIPTION OF THE α GENE

Image removed due to copyright restrictions.

Gene Sequences at a Distance from the Gamma Globin Chain Gene Affect Level of Expression

- Deletions of beta chain gene can lead to increased levels of gamma chain synthesis

Image removed due to copyright restrictions.

Deletions which entirely eliminate the Beta Globin Gene Cause the Gamma Chain Genes to Remain On-- Hereditary Persistence of Fetal Hemoglobin (HPFH)

Worldwide Distribution of Globin Disorders
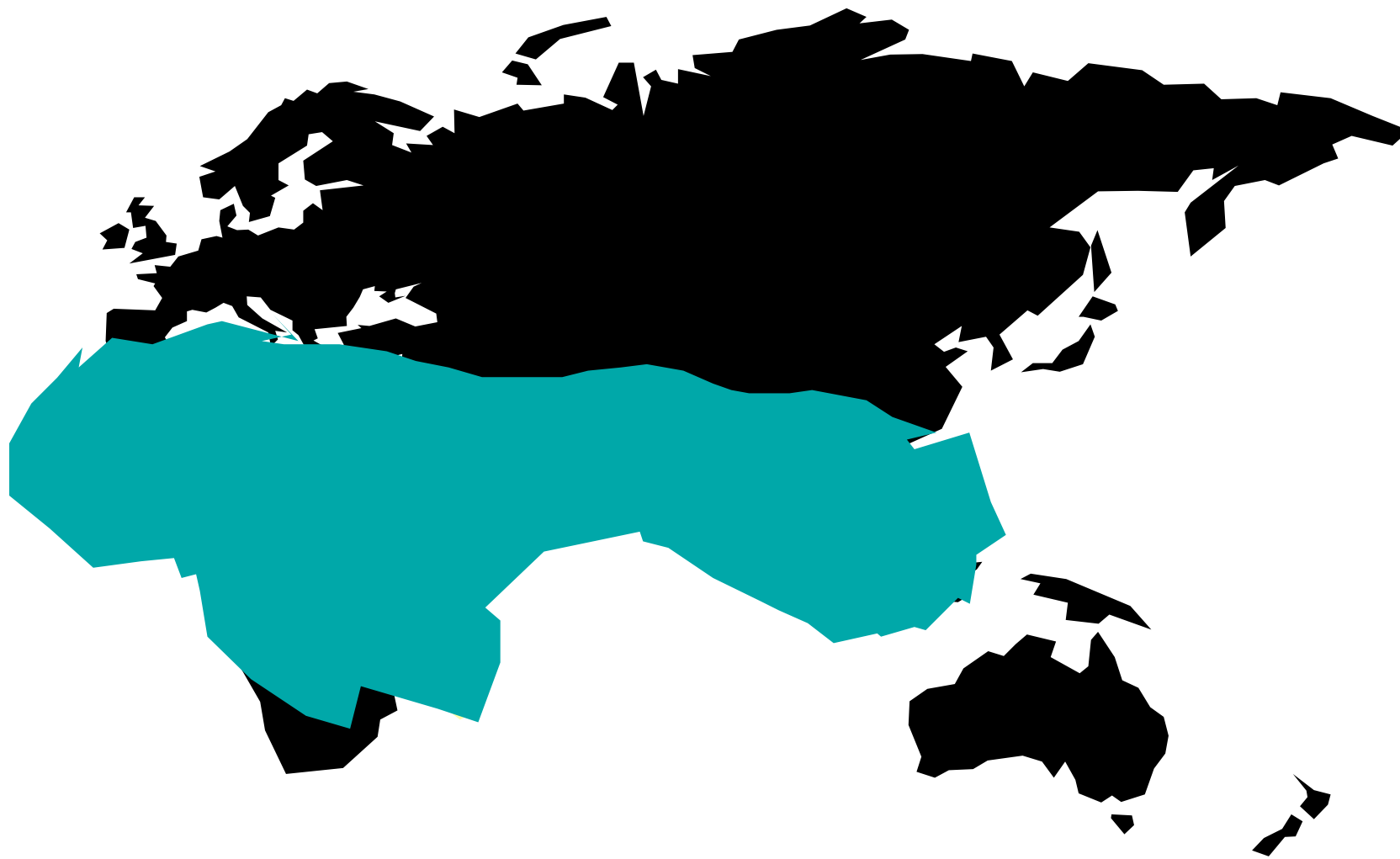
Epidemiology and the malaria hypothesis

Distribution of thalassemias, sickle cell disease, G6PD mirror worldwide distribution of malaria prior to 20th century.

Hypothesis (Haldane and others): heterozygous forms confer fitness - Thal trait, sickle trait, G6PD protective against death from cerebral falciparum malaria
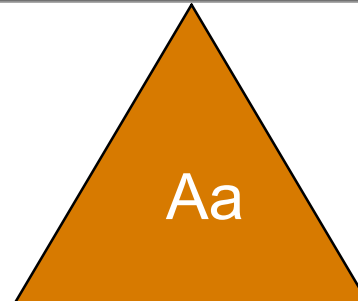
Images removed due to copyright restrictions.
Maps showing the similar worldwide distributions of falciparum malaria and alpha-thalassemia.

Balanced Polymorphism

Malaria
AA

Thalassemia
aa

Aa

Hardy-Weinberg Equilibrium

- large population

- no mutation

- no selection

- random mating

- no migration

[A] = p
[a] = q
p + q =1

[AA] = p^2
[Aa] = 2pq
[aa] = q^2

frequencies remain stable

The only reason that Hardy-Weinberg should not hold at conception is assortative mating

eggs

	A	a	allele
	$\parallel$	$\parallel$	
sperm	p	q	frequency

allele   frequency

A = p

a = q

	A (p)	a (q)
A (p)	AA p^2	Aa pq
a (q)	aA pq	aa q^2

Selection: Genetic Lethal Can Eliminate One Genetic Class at Some Point in Life

	AA	Aa	aa
At conception	p^2	$2pq$	q^2
After lethal events	p^2	$2pq$	0

Selection: Genetic Lethal Can Eliminate One Genetic Class at Some Point in Life

	AA	Aa	aa
At conception	p^2	$2pq$	q^2
Early in life— some lethal events reduce aa class	p^2	$2pq$	Less than q^2
Later in life— more lethal events reduce aa class	p^2	$2pq$	Much less than q^2
After all lethal events	p^2	$2pq$	Can be as low as 0

Balanced Selection: Genetic Lethal Can Eliminate More than One Genetic Class

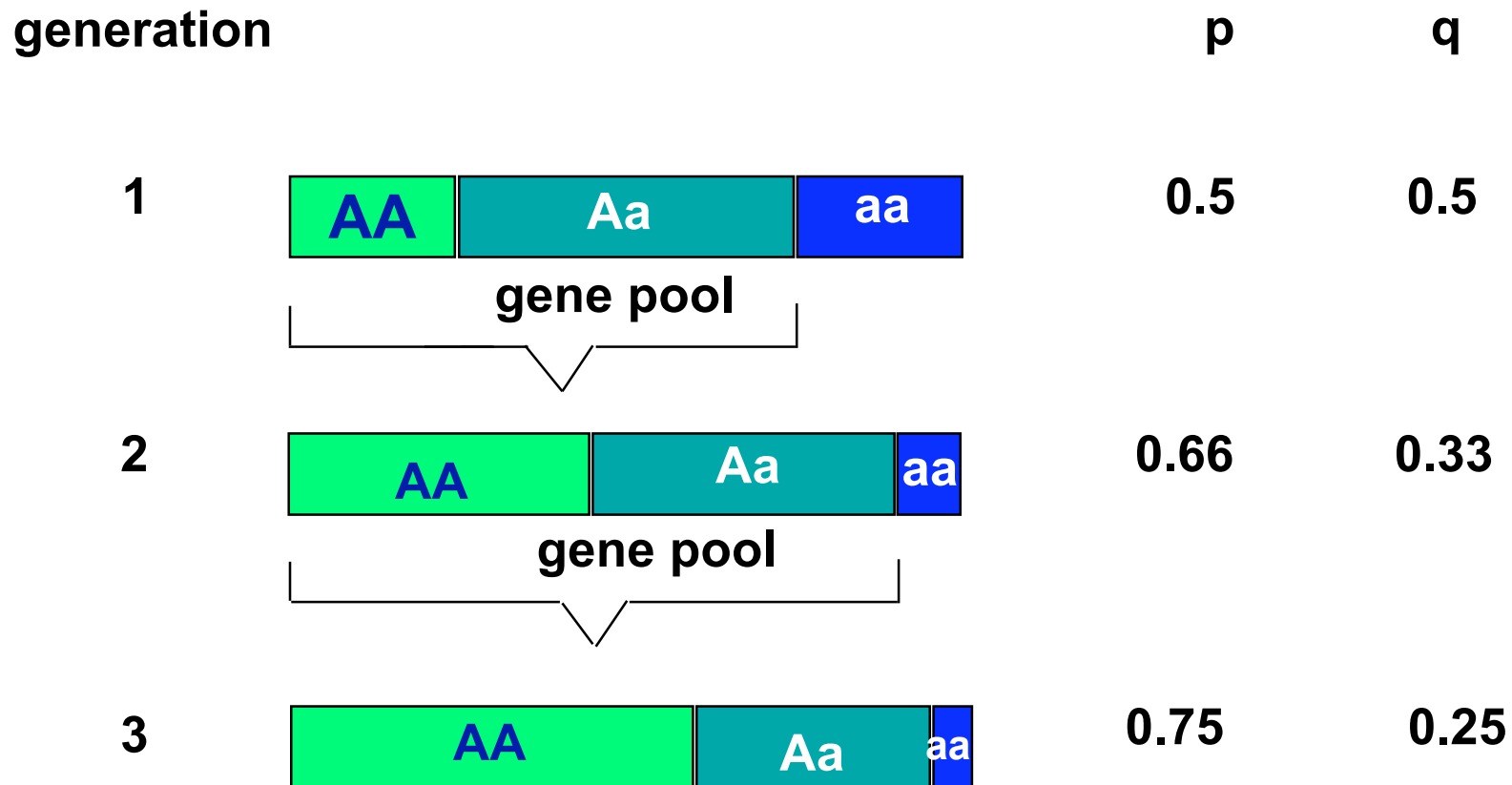|	AA	Aa	aa
At conception	p^2	$2pq$	q^2
Early in life—some lethal events reduce AA and aa classes	Less than p^2	$2pq$	Less than q^2
Later in life—more lethal events reduce AA and aa classes	Much less than p^2	$2pq$	Much less than q^2
After all lethal events	Can be as low as 0	$2pq$	Can be as low as 0

Age Related Allele Frequencies can Demonstrate Genetic Selection Operating on a Population

- Deviations from HWE as a function of age can demonstrate survival effects of alternative alleles at a locus

Genetic Selection Operating on a Population will Influence Allele Frequencies in Subsequent Generations

- Reproductive fitness of each genotype class at the time of mating is the critical parameter in determining the genotype composition of the next generation

What matters for the next generation is the proportion of individuals in each genotype class at the time of mating and their relative effectiveness in mating--REPRODUCTIVE FITNESS

Fitness

fitness: proportion of offspring compared with "normal"

coefficient of selection = 1-F

F = 1, s = 0 if normal number of offspring
F = 0, s = 1 if lethal

# Change in Allele Frequency Because of Reduced Reproductive Fitness

	AA	Aa	aa
At birth	p^2	$2pq$	q^2
Fitness	1	1	$1-s$
In Next Generation	p^2	$2pq$	$q^2(1-s)$

Even though selection may remove all homozygous recessives from the reproductive pool, heterozygote matings recreate this genotype class in the next generation
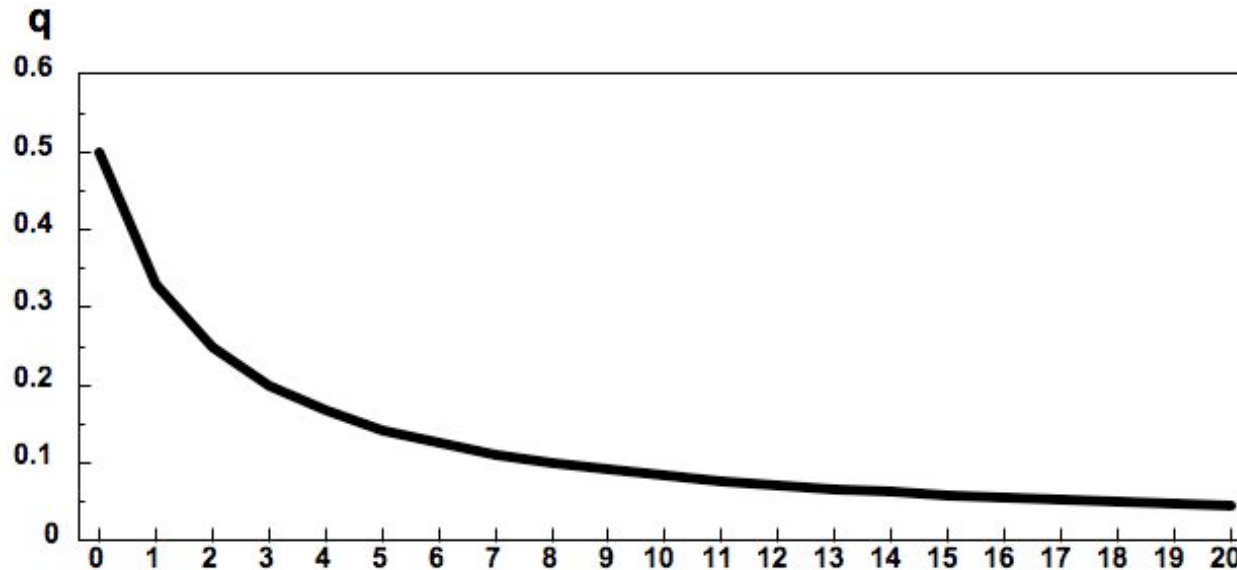
generation		p	q
1	AA Aa aa	0.5	0.5
	gene pool		
2	AA Aa aa	0.66	0.33
	gene pool		
3	AA Aa aa	0.75	0.25
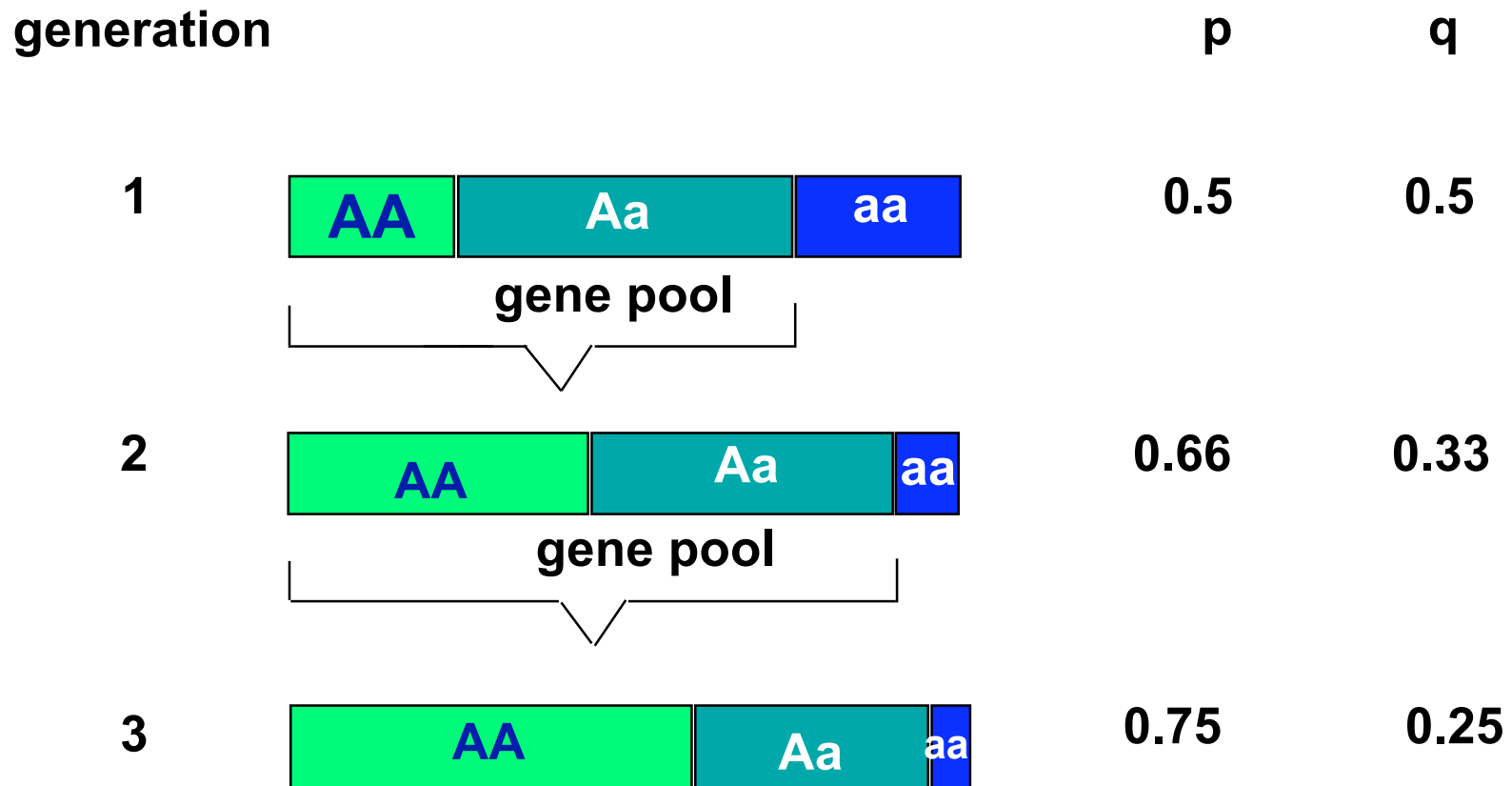
Change in Gene Frequency
with Extremely Reduced Reproductive Fitness of the
Homozygous Recessive
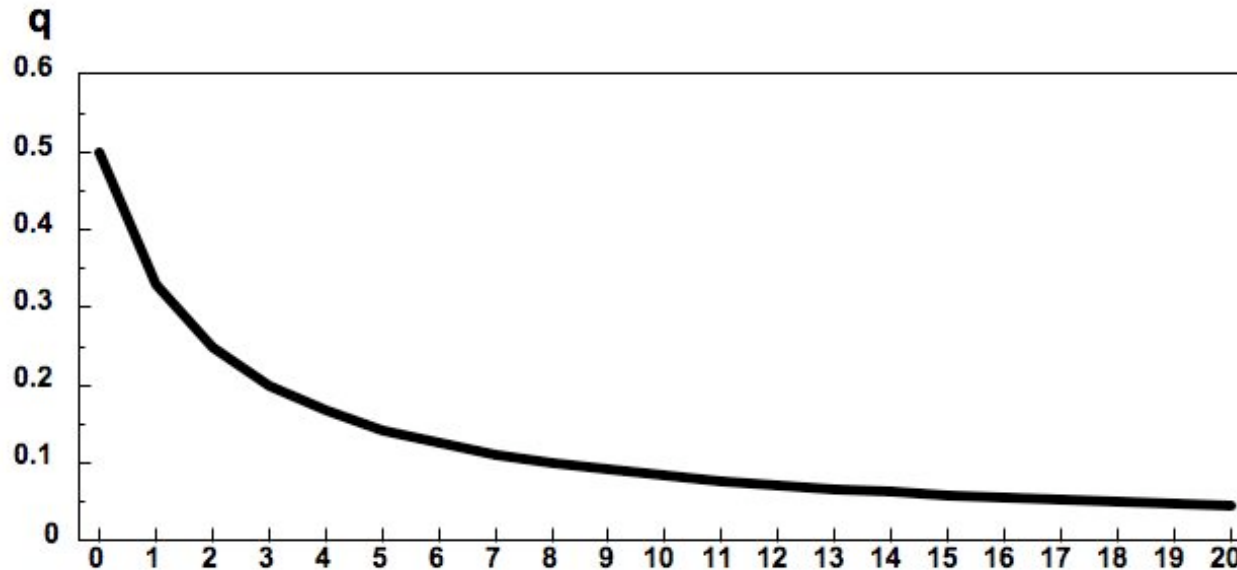
Even after many generations the recessive allele remains at a
significant frequency in the population but over time it gradually
will drop to lower and lower population frequencies UNLESS

Heterozygote Advantage

When allele frequencies for the recessive allele remain high even after many generations despite negative selection of the recessive homozyogote, then selection in favor of the heterozygote should be considered

Examples: sickle cell anemia, thalassemia, cystic fibrosis, hereditary hemochromatosis, ccr5 deletion

Epidemiology and the malaria hypothesis

Homozygote loss of fitness balanced by increased fitness of heterozygote

>200 million people are heterozygous carriers of thal mutations

~1 million with HbE/β^0 thalassemia

tens of thousands with homozygous thal syndromes. syndromes.

Images removed due to copyright restrictions.
Maps showing the similar worldwide distributions of falciparum malaria and alpha-thalassemia.

Image removed due to copyright restrictions.

Graph showing the leading infectious killers in 1998, worldwide: acute respiratory diseases (3.5 million), AIDS (2.3 million), diarrheal diseases (2.2 million), malaria (1.1 million), and measles (0.9 million).

Malaria--Worldwide Impact

- **40% of the world's population** - mostly those living in the poorest countries - is at risk of malaria.
- Malaria causes between **300 - 500 million cases of acute illness** and over **1 million deaths** annually.
- **90% of deaths due to malaria occur in Africa**, south of the Sahara desert - mostly amongst young children.
- Malaria is the number one killer of young children in Africa, accounting for **1 in 5 of all childhood deaths**

Species of Plasmodium which cause Malaria

- P. falciparum is the major killer among the four species of plasmodium which infect humans

- P. vivax, p. malariae and p.ovale cause serious illness but are not responsible for the high death rate caused by malaria in humans

Plasmodium falciparum has a very broad world wide distribution. It has a tertain pattern of fevers. It is responsible for most malaria fatalities. Infected red blood cells develop surface 'knobs' which cause them to stick to endothelial cells (cells lining blood vessels). This causes blockages and brain and intestinal damage, often resulting in death, which can occur within a few days of infection. *P. falciparum* is especially dangerous to small children and to travellers from non-malarious areas. There is no dormant hyponozoite stage.

Image removed due to copyright restrictions. Life stages of P. falciparum.

What about immunity?

- Adults from areas where malaria is endemic develop a form of partial   immunity.
- This partial immunity develops slowly and only in response to repeated infections.
- In partially immune people, malaria parasites can often be found in the blood, but without clinical symptoms.
- Immunity is lost if exposure is not maintained. (after 6 months).

Malaria Pathology

- P. falciparum can cause repeated inefections; immune response is very limited
- P. falciparum causes infected red blood cells to exhibit protruding knobs which stick to endothelial cells lining blood vessels
- Blockage of blood vessels in the brain causes cerebral malaria--a major cause of death
- Blockage of intestinal blood vessels also important in pathology

How the malaria life cycle in mosquitos determines where malaria will be most prevalent
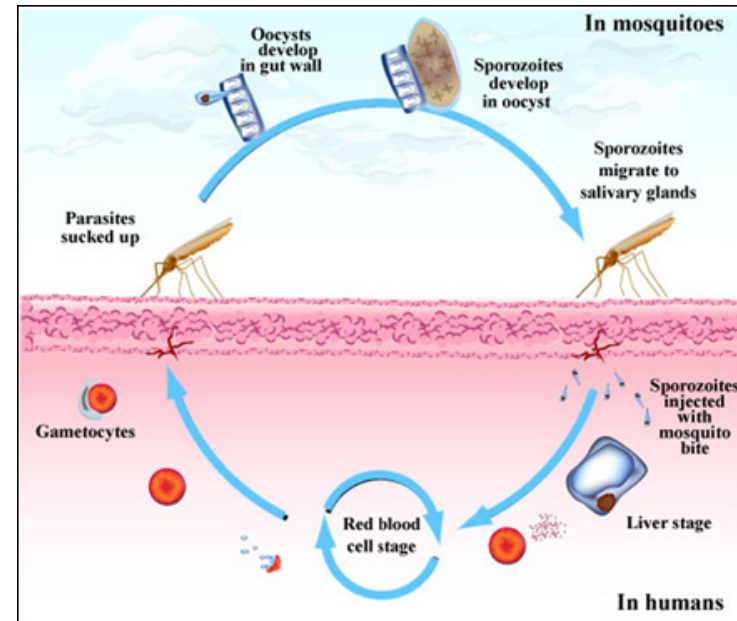
In the part of the life cycle carried out in mosquitos, the embedded ookinete becomes an oocyst which grows rapidly and divides internally into sporozoites - the third asexual phase. The oocyst is the longest phase in the life cycle lasting between 8 and 35 days. How long this phase takes is acutely dependent on temperature.

The mosquito has to survive long enough for the oocyst to mature before it can infect anyone. So only elderly mosquitoes can pass on malaria and in the wild of course many mosquitoes never reach old age.

This is where temperature matters, the warmer the weather, the faster the oocyst can develop in the mosquito. Mosquito survival is probably the single most important factor in malaria transmission.

Demonstrating the protective effect of red cell mutations against p. falciparum infection

	In vitro		In vivo Case-control	References
	Invasion/Growth	Phagocytosis		
HbAS	↓ invasion/multiplication (low O_2 conditions)	↑ susceptibility to phagocytosis	Protection: severe malaria (60-90%); mortality (55%)	Willcox et al. (1983a); Aidoo et al. (2002); Ayi et al. (2004)
HbAC	normal	N/A	Reduced risk: clinical malaria (29%); severe malaria (47-80%)	Friedman et al. (1979); Agarwal et al. (2000); Modiano et al. (2001)
HbCC	↓ multiplication; altered knob formation	↑ susceptibility to phagocytosis	Reduced risk of clinical malaria (90%)	Fairhurst et al. (2003); Mockenhaupt et al. (2004a)
HbAE	↓ parasite invasion (25%)	↑ susceptibility to phagocytosis	Reduced risk of complications (6.9x reduced odds)	Yuthavong et al. (1990); Hutagalung et al. (1999); Chotivanich et al. (2002)
HbEE	↓ invasion/multiplication	↑ susceptibility to phagocytosis	N/A	
α-Thalassaemia	Normal invasion ↓ multiplication	No difference from controls	Reduced risk of severe malaria (risk = 0.4 - 0.66)	Allen et al. (1997); Pattanapanyasat et al. (1999); Mockenhaupt et al. (2004b)
β-Thalassaemia	inconclusive	↑ susceptibility to phagocytosis	Protection against hospital admission with malaria (50%)	Willcox et al. (1983b); Ayi et al. (2004);
G6PD deficiency	↓ growth under oxidative stress	↑ susceptibility to phagocytosis (ring stage)	Protection of F(htz) against non-severe malaria; Protection of M and F(htz) against severe malaria (50%)	Gilles et al. (1967); Bienzle et al. (1972); Ruwende et al. (1995); Cappadoro et al. (1998)
PK deficiency	Unknown	Unknown	Unknown	Min-Oo et al. (2003);

The observed malaria protective effect of common erythrocyte variants.

↓, decrease; ↑, increase; M, male; F, female.

Figure by MIT OpenCourseWare. After Min-Oo and Gros, 2005.

High Frequency of Alpha Thalassemia Alleles in Regions of the World with High Incidence of Falciparum Malaria

Image removed due to copyright restrictions.
Map of global incidence of alpha-thalassemia alleles.

High Frequency of Beta Thalassemia Alleles in Regions of the World with High Incidence of Falciparum Malaria

Image removed due to copyright restrictions.
Map of global incidence of beta-thalassemia alleles.

Different βThalassemia mutations are most common in different populations

| Geographic area	Description
Mediterranean	IVS-1, position 110 (G → A) Codon 39, nonsense (CAG → TAG) IVS-1, position 1 (G → A) IVS-2, position 745 (C → G) IVS-1, position 6 (T → C) IVS-2, position 1 (G → A)
African	−29, (A → G) −88, (C → T) Poly(A), (AATAAA → AACAAA)
Southeast Asian	Codons 41/42, frameshift (−CTTT) IVS-2, position 654 (C → T)
Indian sub-continent	−28, (A → G) IVS-1, position 5 (G → C) 619-bp deletion Codons 8/9, frameshift (+G) Codons 41/42, frameshift (−CTTT) IVS-1, position 1 (G → T)

Why do different human populations have different allele frequencies for many genetic loci?

- Genetic drift
- Founder effects
- Mutation
- Selection

Genetic Drift

- fluctuation in gene frequency due to small size of breeding population
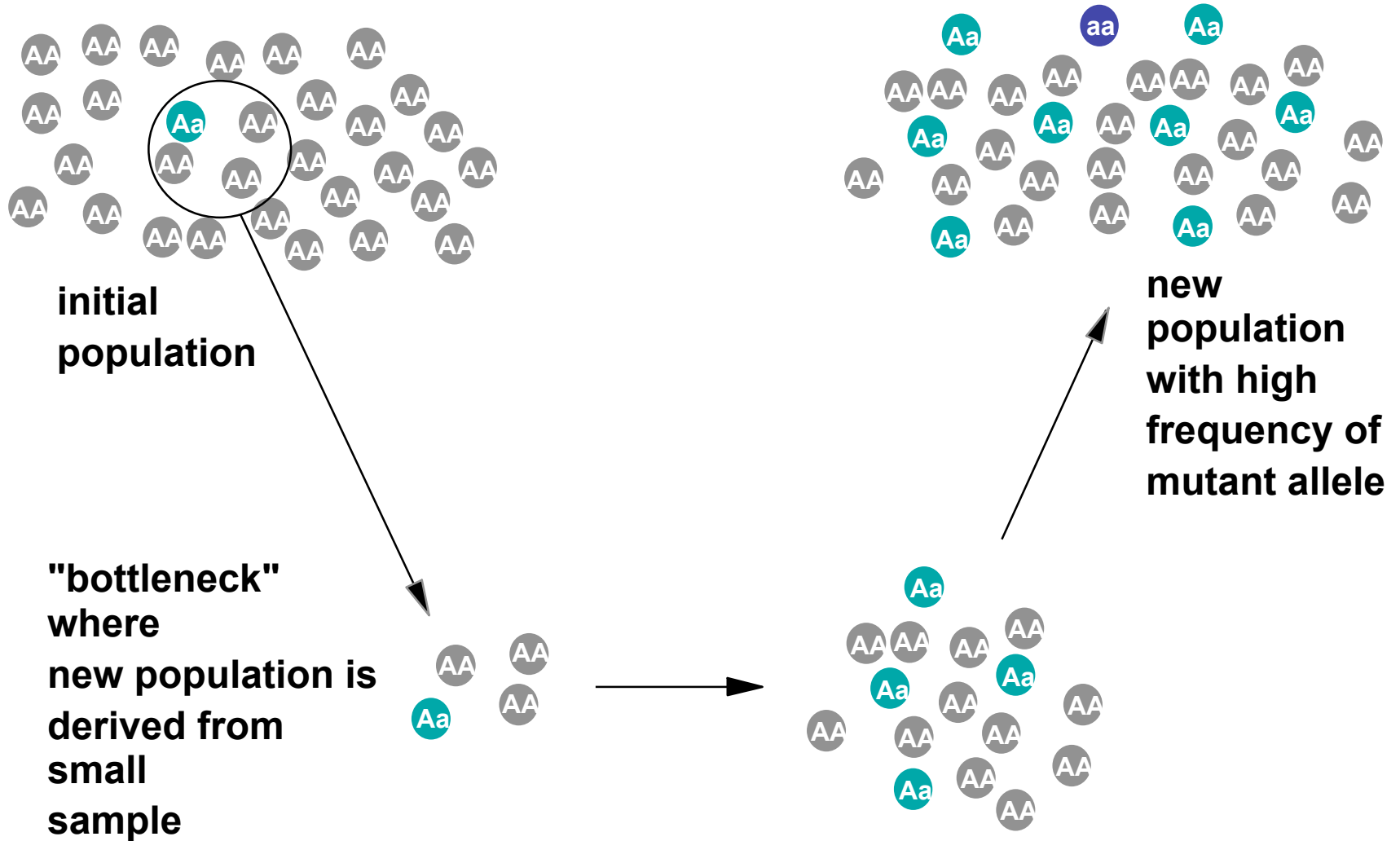- fixation or extinction of allele possible

Genetic Drift
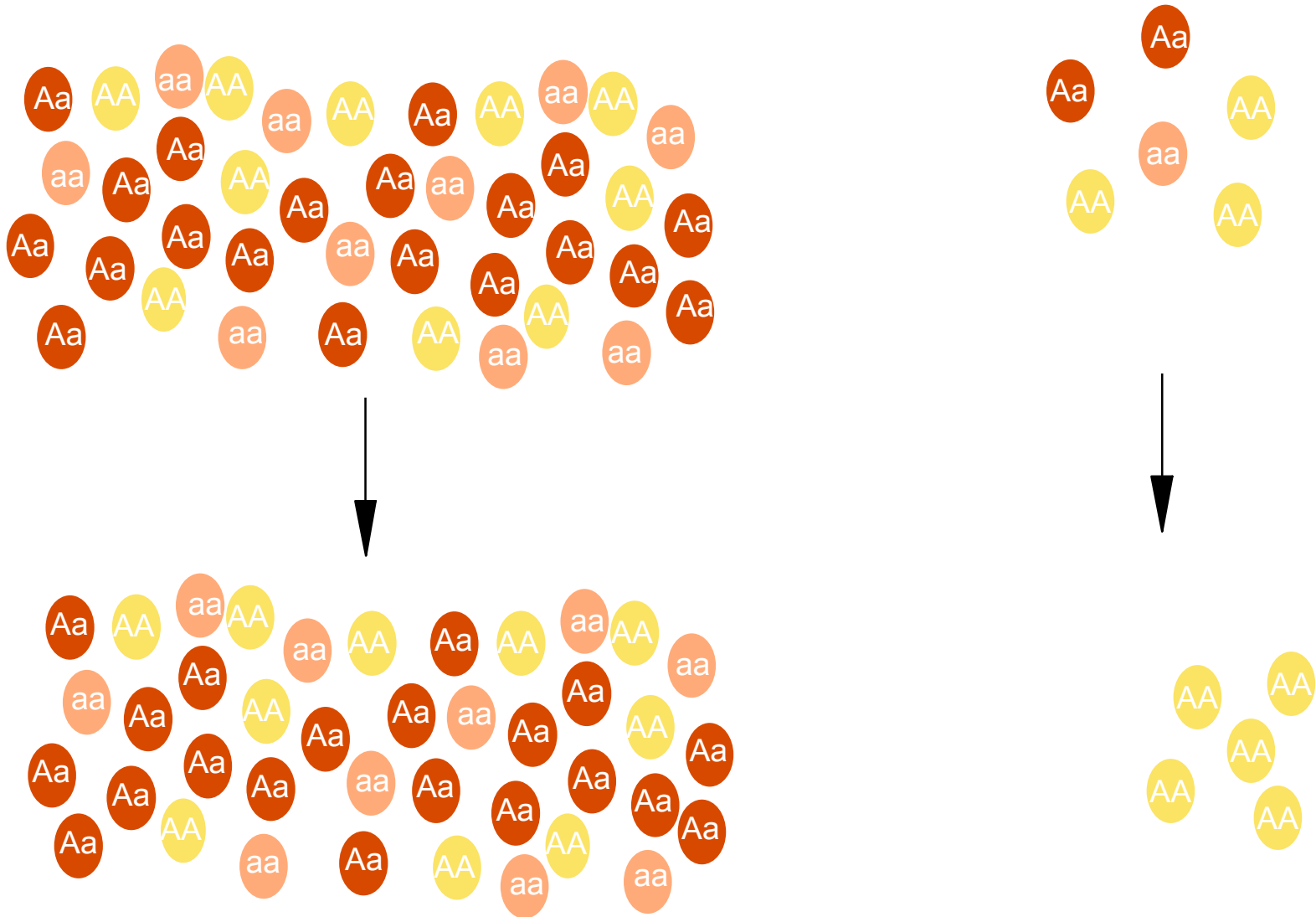
Founder Effect

- high frequency of gene in distinct population
- introduction at time when population is small
- continued relatively high frequency due to population being "closed"

Founder Effect

initial
population

"bottleneck"
where
new population is
derived from
small
sample

new
population
with high
frequency of
mutant allele

POSITIONAL CLONING

- RESOLUTION IN FAMILY ANALYSIS IS LIMITED BY THE NUMBER OF FAMILY MEMBERS YOU CAN COLLECT

- BUT IF THE AFFECTED INDIVIDUALS IN A POPULATION CAN BE TREATED ESSENTIALLY LIKE A LARGE FAMILY, THEN INCREASED RESOLUTION CAN BE ACHIEVED

What if you want better resolution than the families collected can give?

- Linkage disequilibrium can sometimes be used to pinpoint position of disease gene
- Linkage disequilibrium uses meioses that happened many generations in the past to increase the resolution of genetic mapping

Linkage Disequilibrium Can Be Used to Identify the Position on a Chromosome Most Likely to Contain the Disease Gene Because of a Shared Chromosomal Haplotype Among Affected Individuals
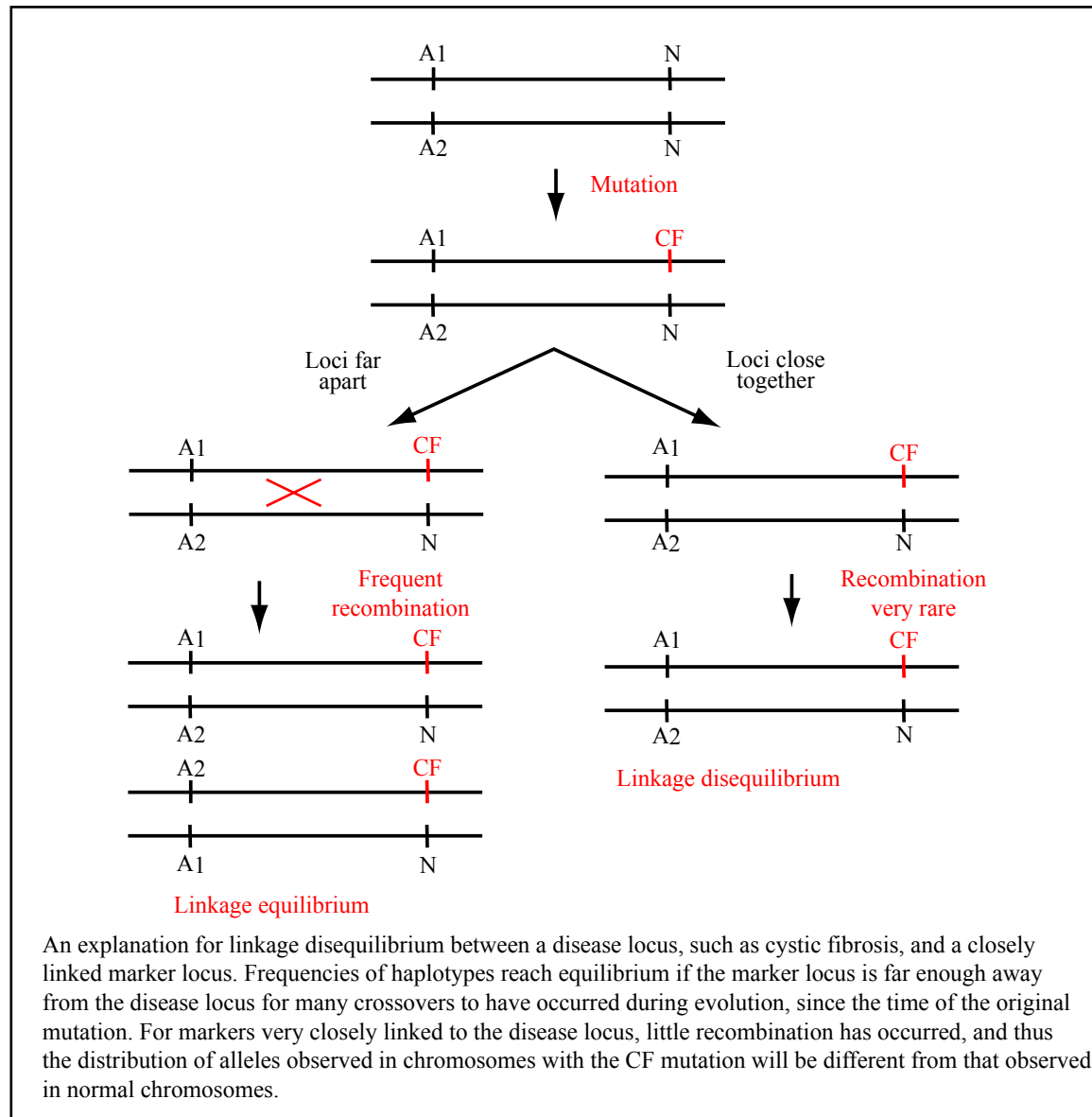

An explanation for linkage disequilibrium between a disease locus, such as cystic fibrosis, and a closely linked marker locus. Frequencies of haplotypes reach equilibrium if the marker locus is far enough away from the disease locus for many crossovers to have occurred during evolution, since the time of the original mutation. For markers very closely linked to the disease locus, little recombination has occurred, and thus the distribution of alleles observed in chromosomes with the CF mutation will be different from that observed in normal chromosomes.
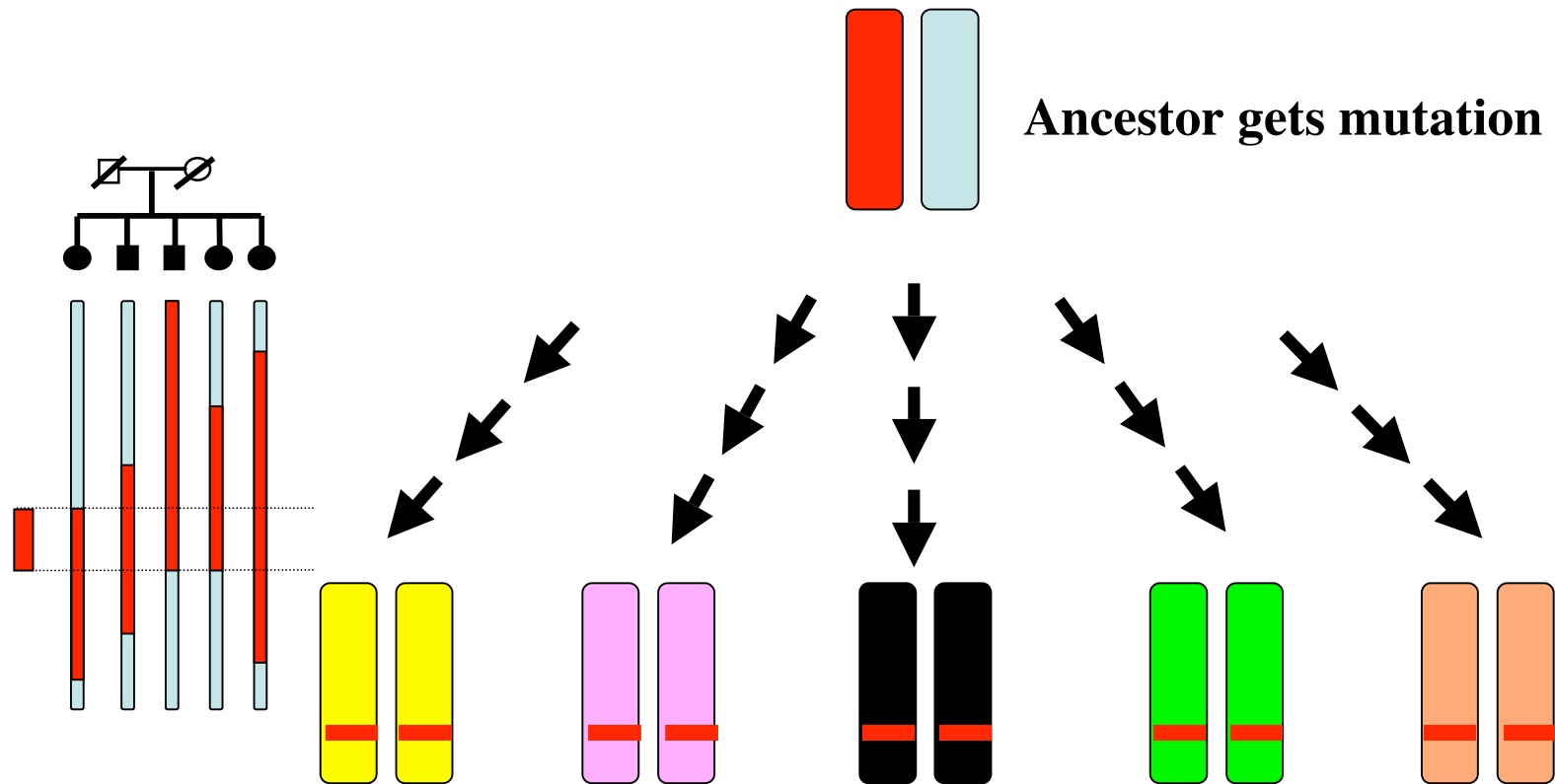
Haplotype

An associated set of alleles at two or more genetic loci on a chromosome which are inherited together over multiple generations because the recombination rate between them is very low relative to the number of generations considered

Linkage disequilibrium

If the alleles at a pair of loci are looked at jointly in a large number of individuals, the presence of a particular allele at one locus may be correlated with the presence of a particular allele at the other locus; when this occurs the two loci are in linkage disequilibrium. One major reason for two loci to show linkage disequilibrium is that the two loci are very close together on a chromosome and the mutation to give rise to one of the alleles has happened recently enough that genetic recombination has not yet caused the new mutant allele to be associated with the two alleles at the other locus with equal frequency
 i.e. they form a haplotype

The principle of positional cloning--affected individuals in a family share a chromosomal segment which includes the mutation causing the disease--Supposing many patients are actually related to each other even though we do not know this for certain


Ancestor gets mutation

**Descendants with disease allele carry region of identity to ancestor and to each other**

Linkage disequilibrium and haplotype analysis can be used to identify location of disease gene if many individuals with disease are actually related to a common ancestor

THE RELATIONSHIP BETWEEN DNA VARIANTS ON A CHROMOSOME AND THE FUNCTIONAL "DISEASE ALLELE" ARE A PRODUCT OF THE HISTORY OF MUTATIONS TO CREATE THE CHROMSOMAL HAPLOTYPE

Emergence of Variations Over Time
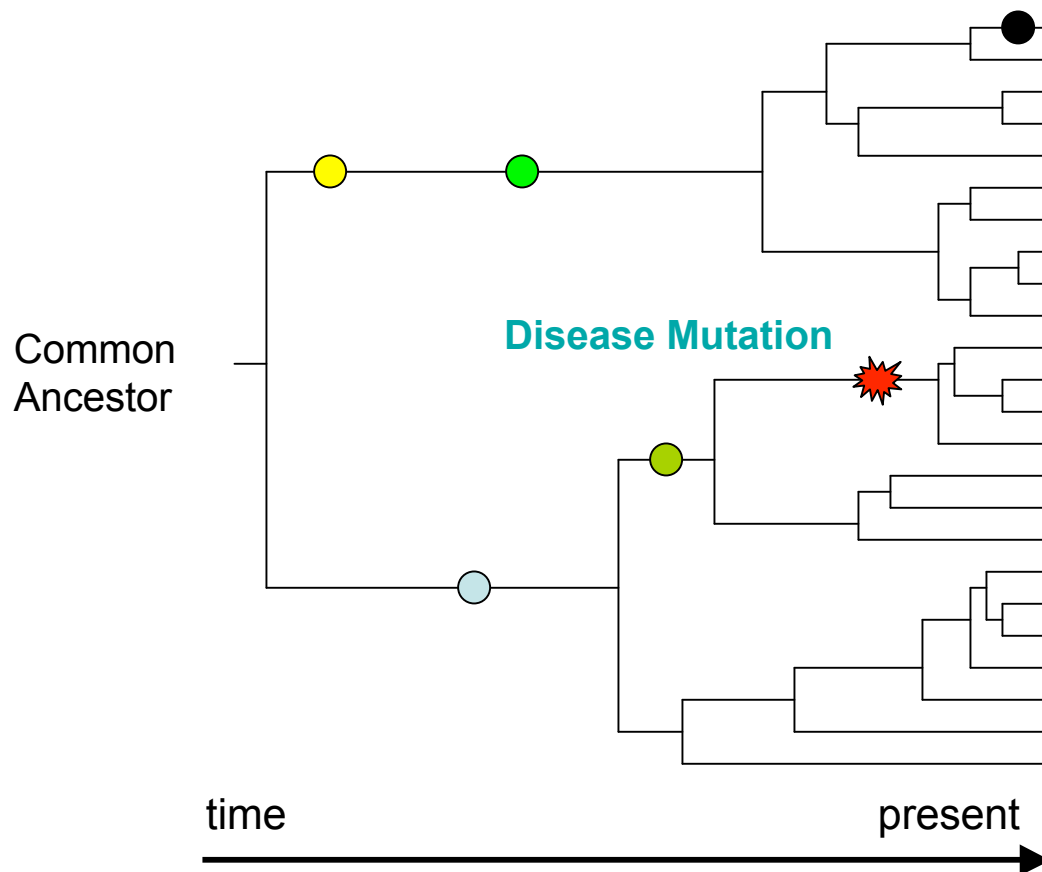
Variations in Chromosomes Within a Population

Common Ancestor

Disease Mutation

time

present
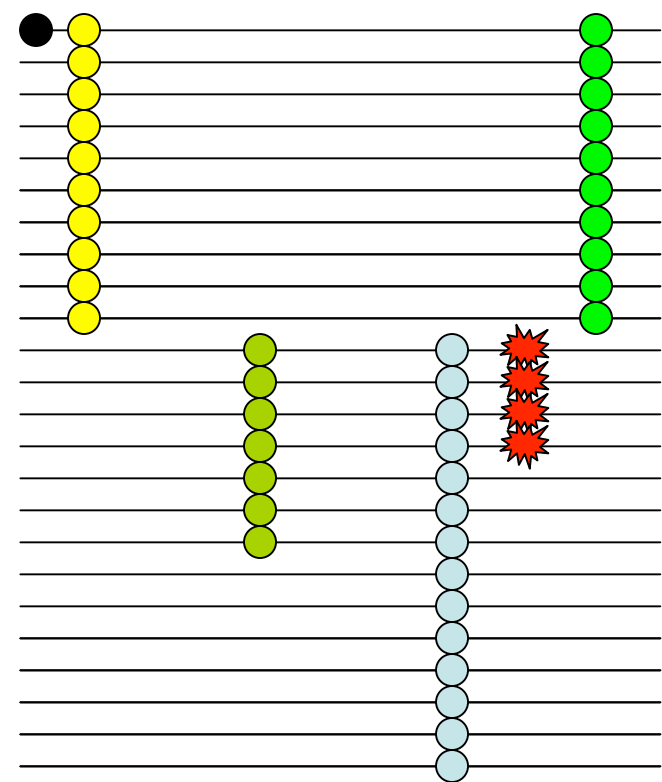
Linkage Disequilibrium Between Marker KM 19 and cystic fibrosis

	ALLELE 1	ALLELE 2
CF CHROMOSOMES	24	228
NORMAL CHROMOSOMES	284	66

Linkage Disequilibrium Between Marker XV2C and cystic fibrosis

	ALLELE 1	ALLELE 2
CF CHROMOSOMES	235	109
NORMAL CHROMOSOMES	17	141