

Neutrality and some of its deviations

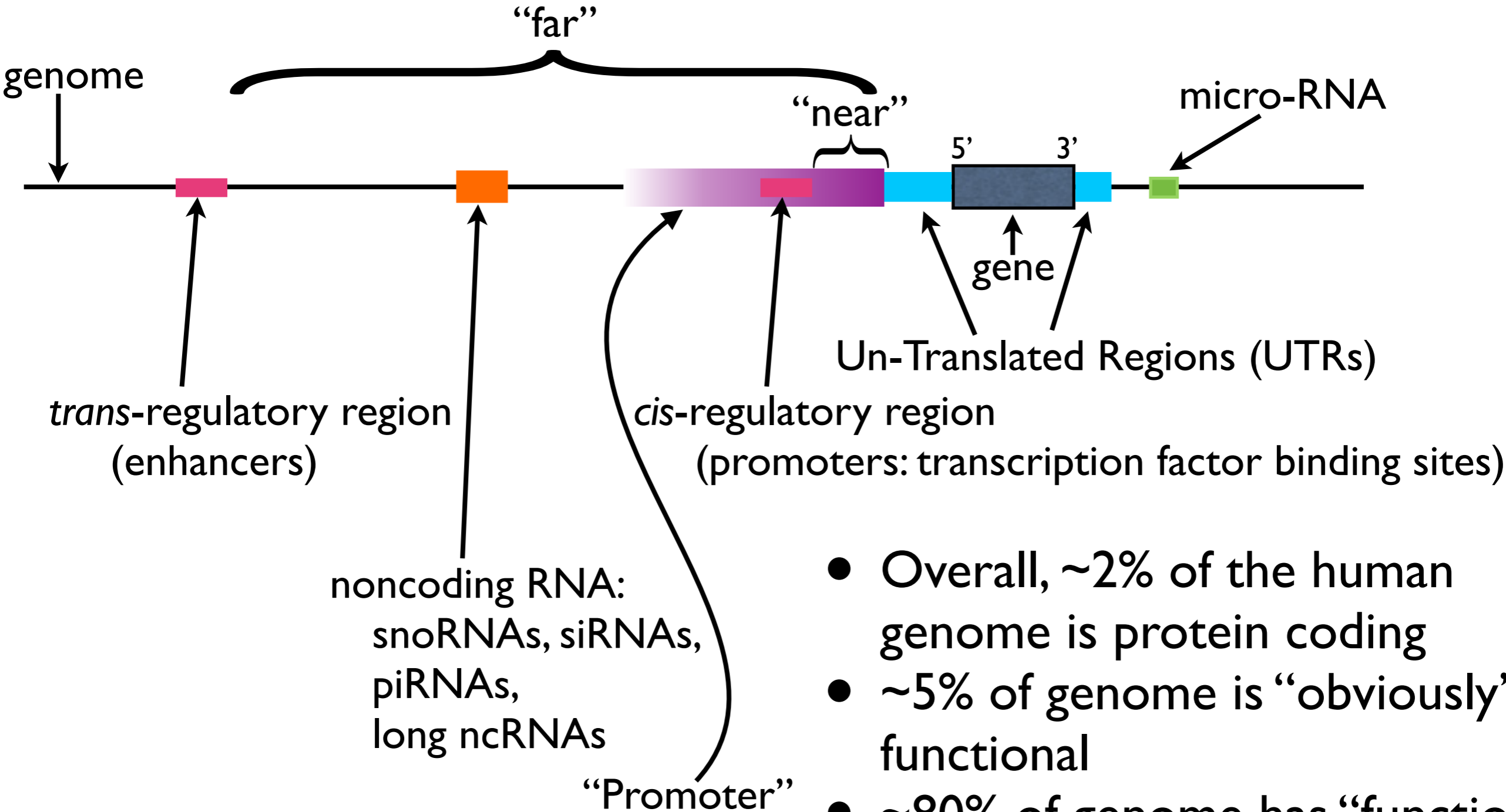
Ryan Hernandez

Goals

- Learn about the population genetics view of the life cycle
- A few Pop Gen summary statistics
- Revisit Hardy-Weinberg Equilibrium - Assumptions & violations

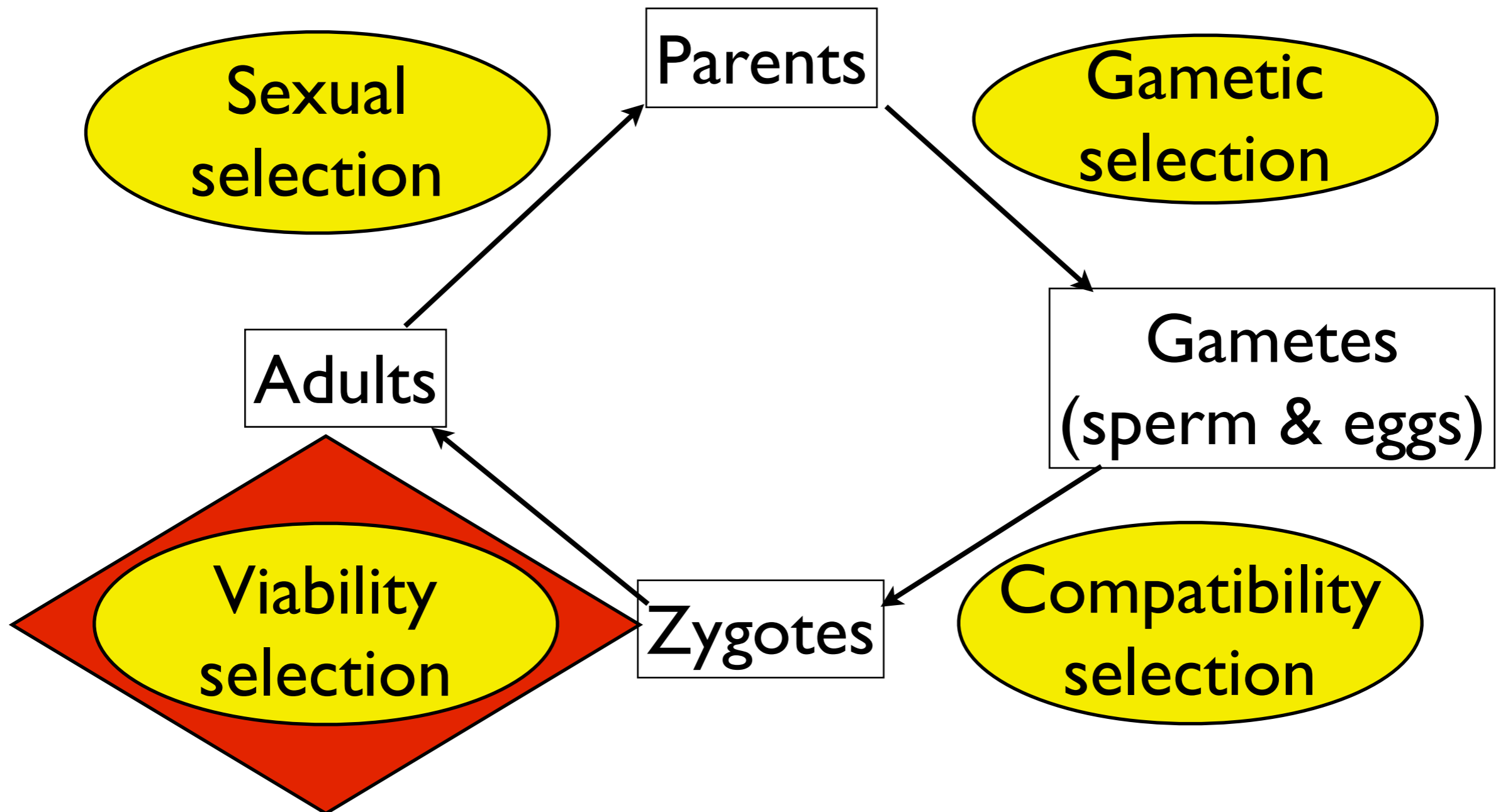
Basic Biology of Human Genome

Functional non-coding mutations



- Overall, ~2% of the human genome is protein coding
- ~5% of genome is “obviously” functional
- ~80% of genome has “functional activity”

Life Cycle

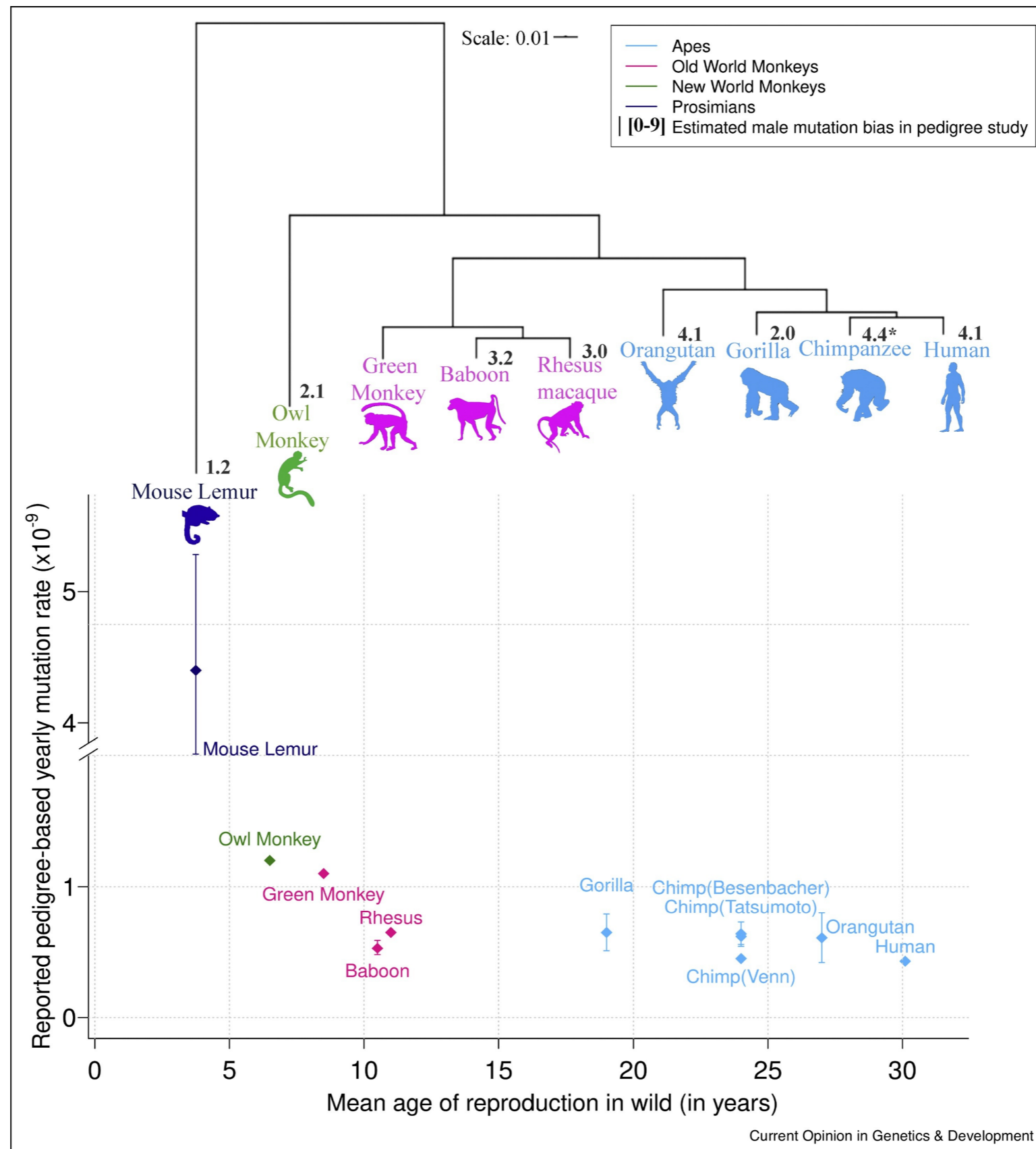


MODERN HUMAN GENOMICS: A CASE FOR RARE VARIANTS?

$$1.1 \times 10^{-8} \times 6 \times 10^9 = 66 \text{ [muts / person]}$$

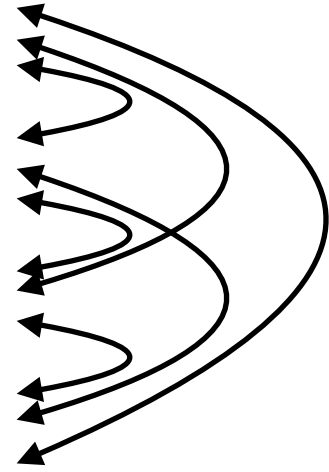
$$\begin{array}{r} 66 \text{ [muts/p]} \\ \times 130\text{M [p/y]} \\ \div 3\text{B [bp]} \\ \hline 2.86 \text{ muts/bp/yr} \end{array}$$

MUTATION RATE EVOLUTION IN PRIMATES



SEQUENCING DATA

Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T

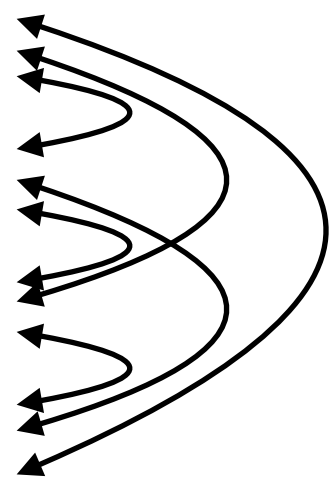


# Pairwise differences	3	4	3	3	3	3
------------------------	---	---	---	---	---	---

π = average pairwise diversity

SEQUENCING DATA

Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T

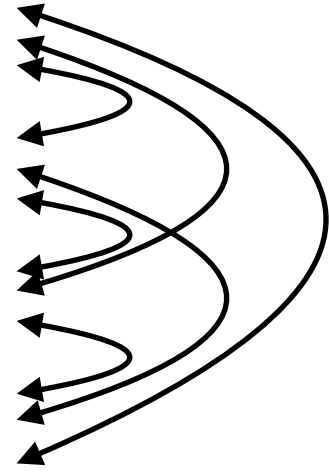


# Pairwise differences	3	4	3	3	3	3
# Compared	6	6	6	6	6	6

π = average pairwise diversity

SEQUENCING DATA

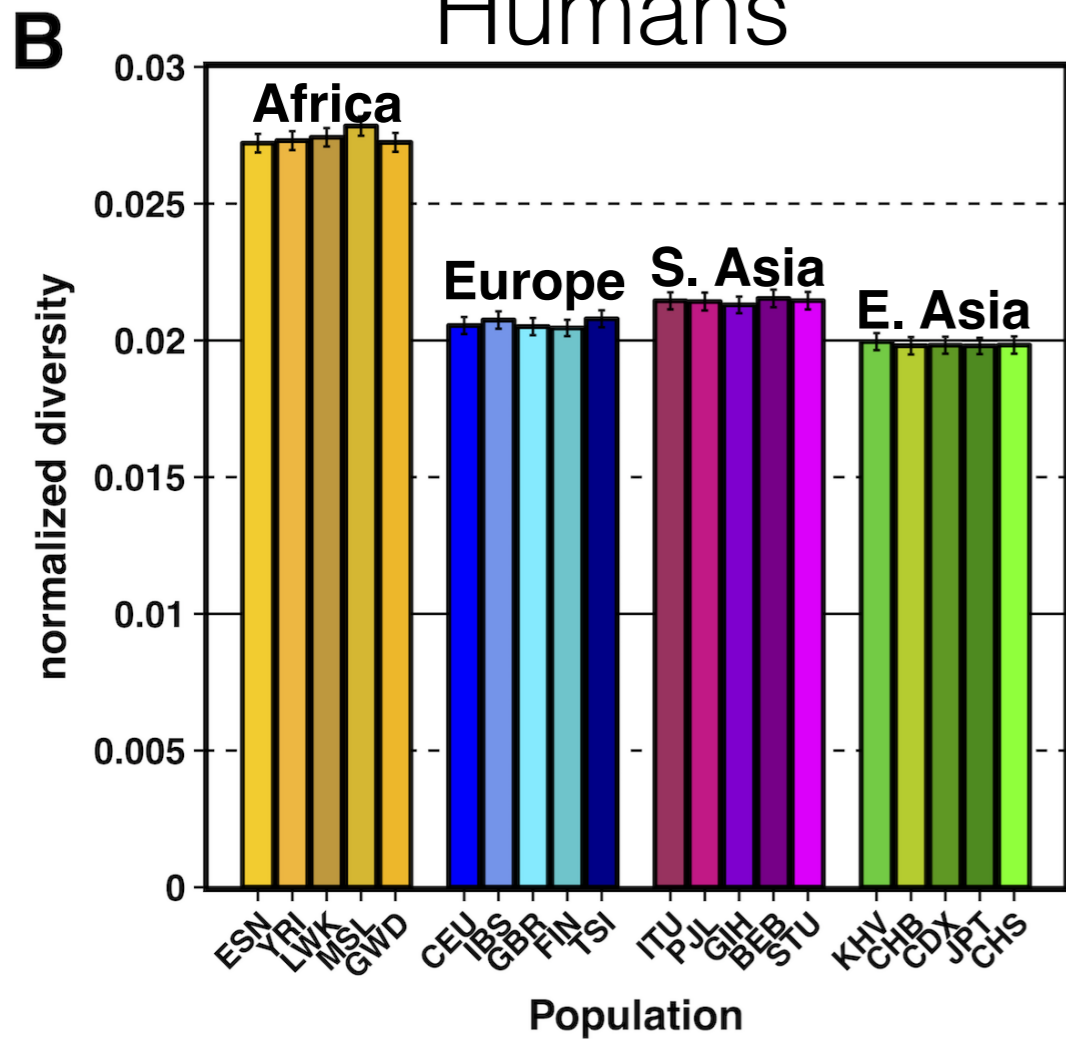
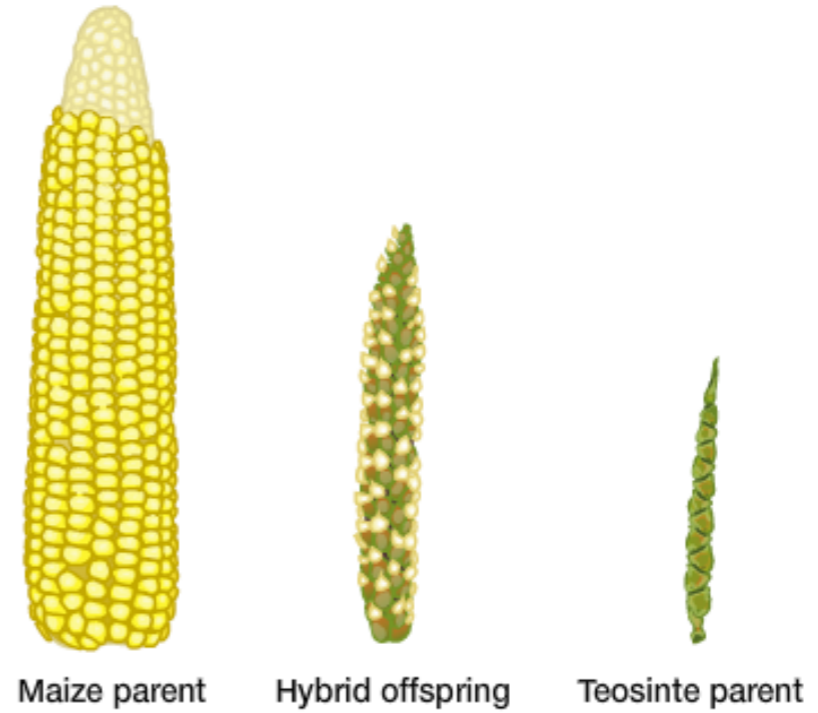
Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T



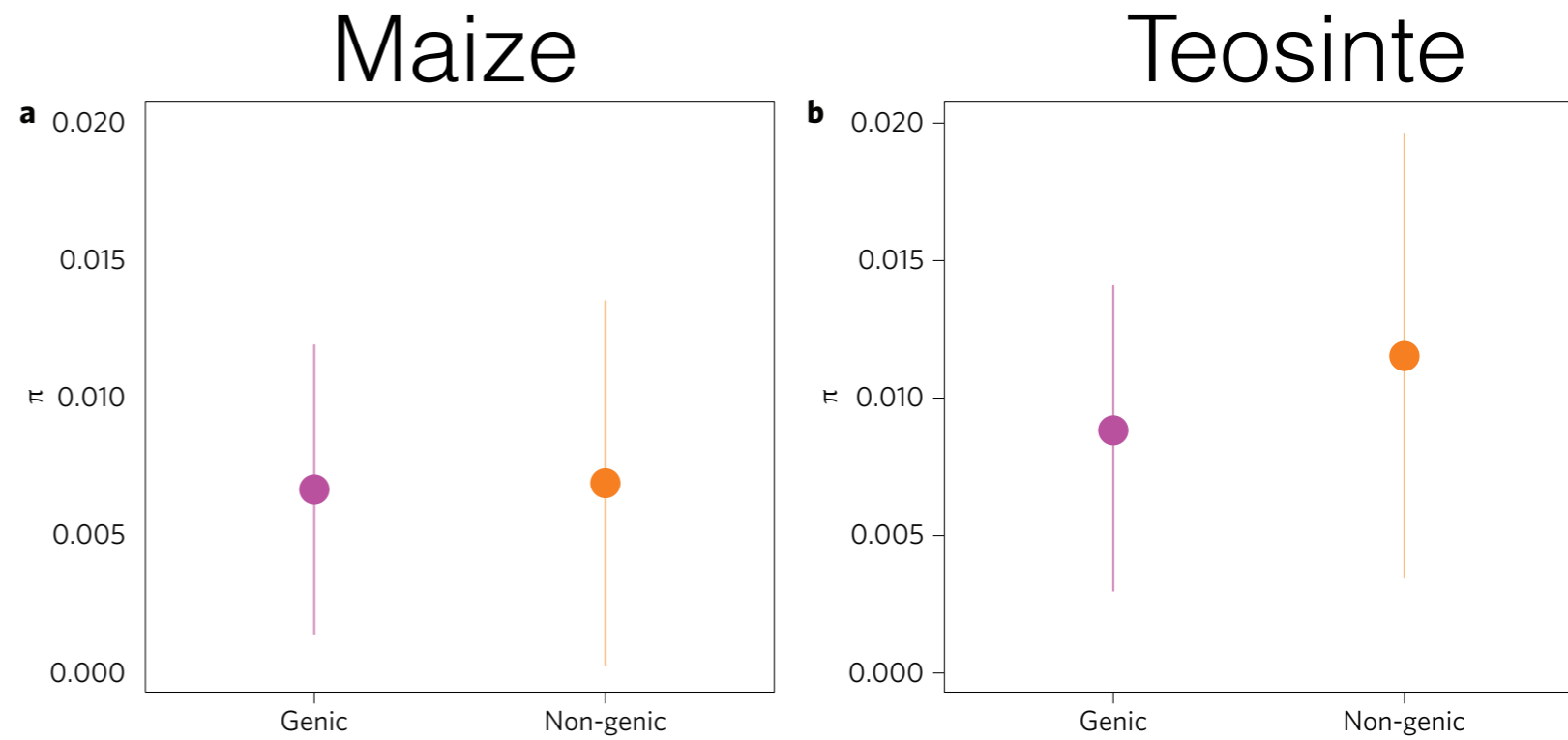
# Pairwise differences	3	4	3	3	3	3
# Compared	6	6	6	6	6	6
Avg. Pairwise Diff	0.5	0.67	0.5	0.5	0.5	0.5

Number of variants: 6 SNPs
 Diversity (π): 3.1667/L

DIVERSITY ACROSS POPULATIONS



Torres, et al. (2018)

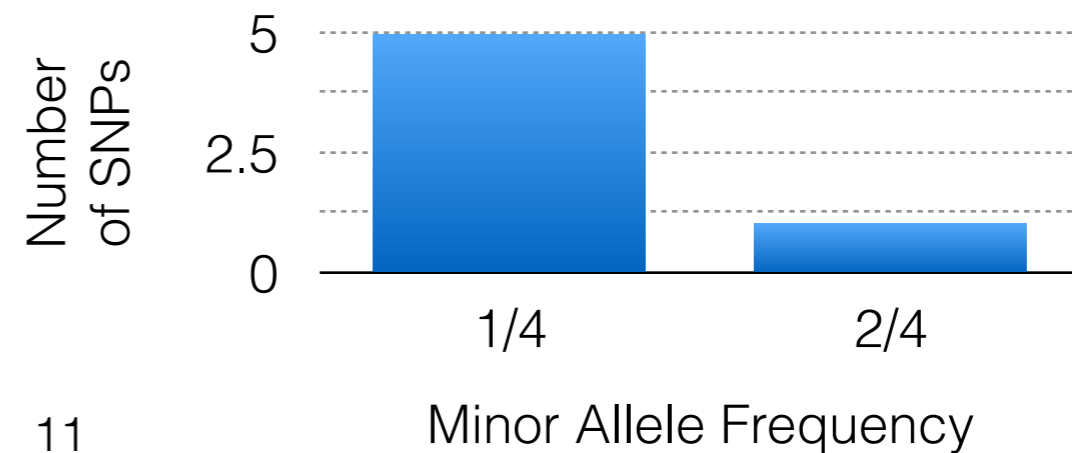


Beissinger, et al. (2016)

SEQUENCING DATA

Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T
Minor Allele	G	T	A	G	T	C
MAF	0.25	0.5	0.25	0.25	0.25	0.25

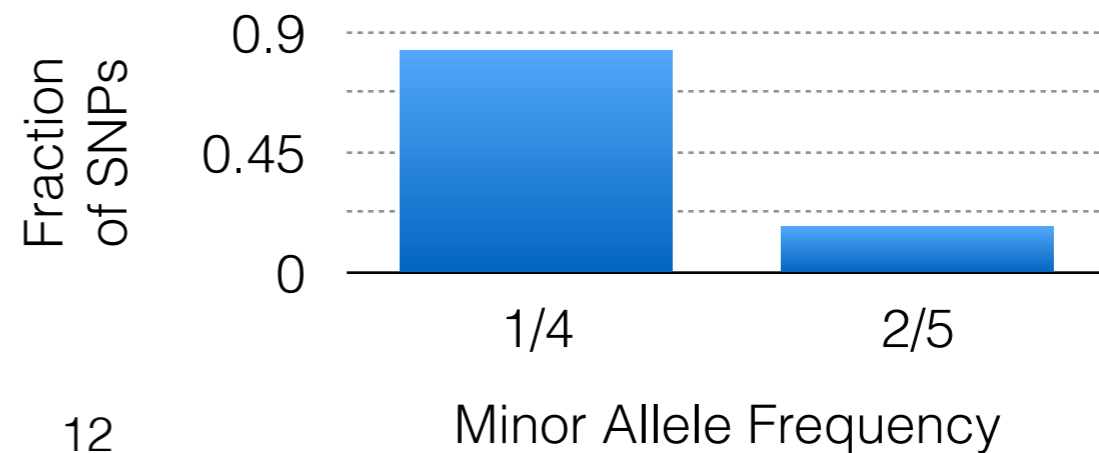
MAF	5	1
-----	---	---



SEQUENCING DATA

Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T
Minor Allele	G	T	A	G	T	C
MAF	0.25	0.5	0.25	0.25	0.25	0.25

MAF	5	1
-----	---	---



SEQUENCING DATA

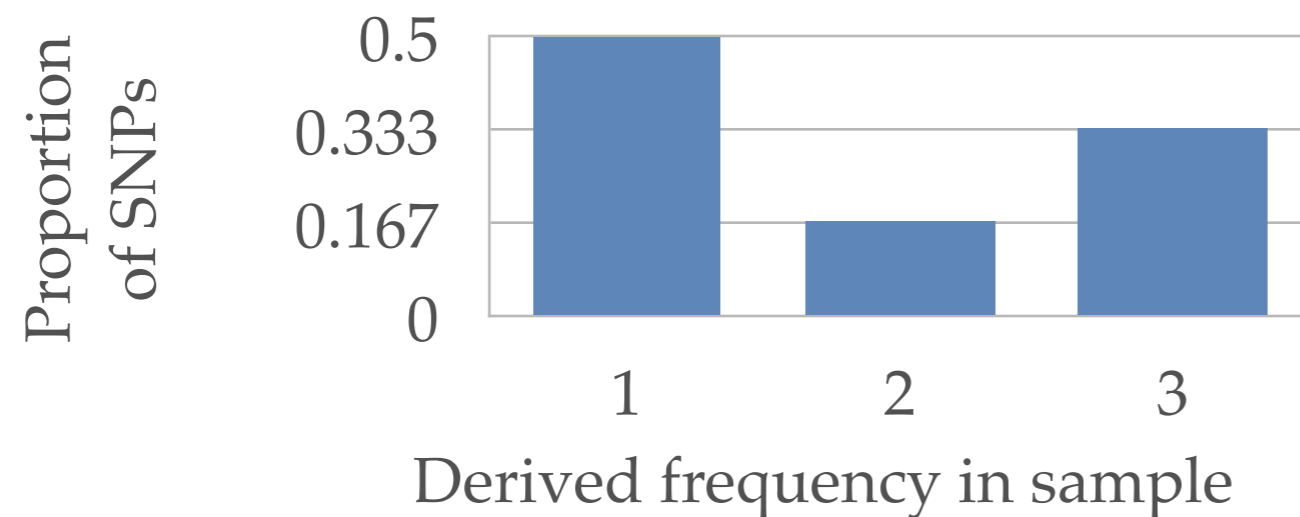
Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T
Chimp	A	C	A	G	C	T

SEQUENCING DATA

Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T
Chimp	A	C	A	G	C	T

SEQUENCING DATA

Chromosome	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
1	A	C	A	G	C	C
2	A	T	G	A	C	T
3	G	T	G	A	T	T
4	A	C	G	A	C	T
Chimp	A	C	A	G	C	T
Derived count	1	2	3	3	1	1



Site-Frequency Spectrum (SFS)

Site-Frequency Spectrum

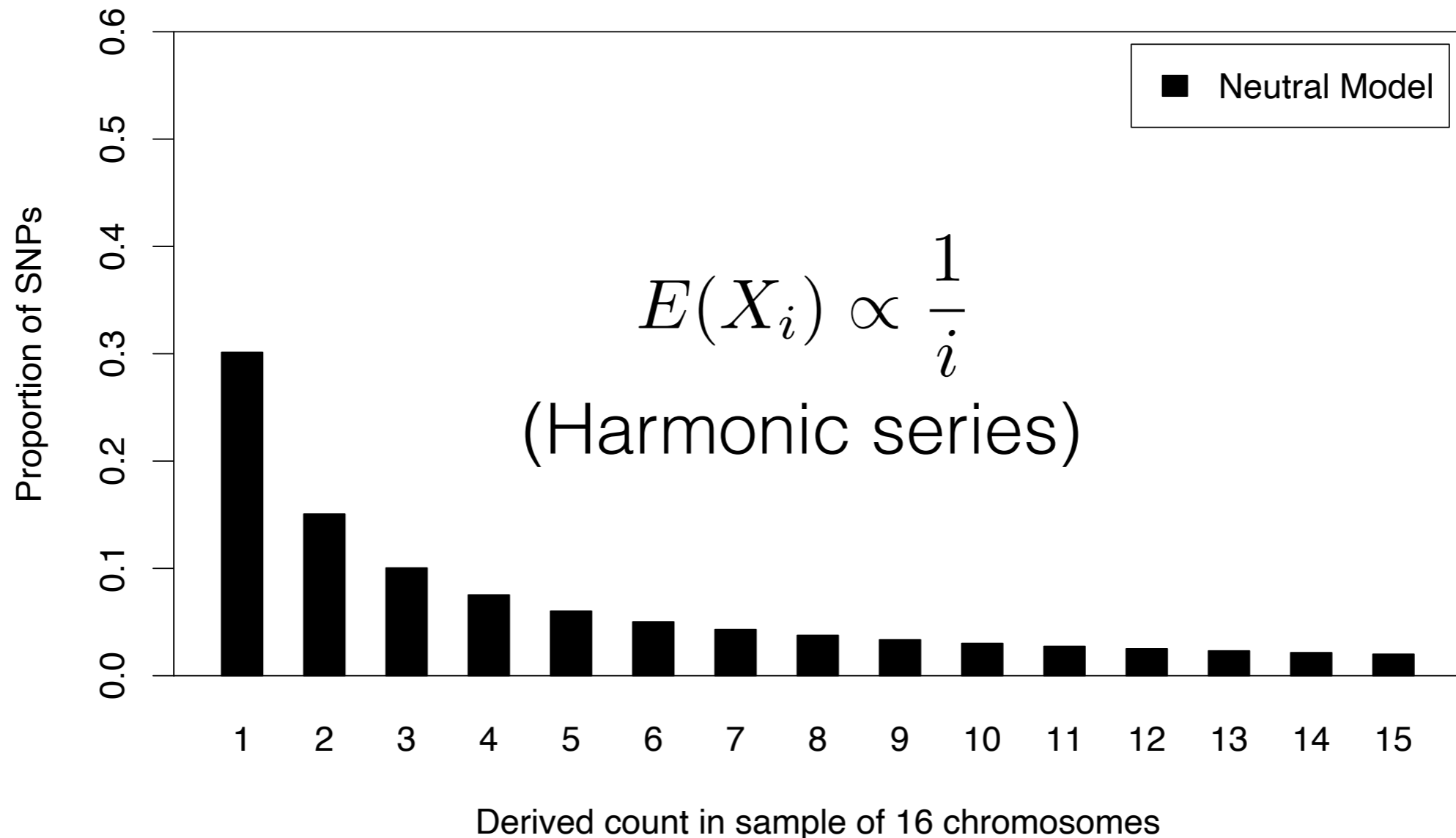
		*		*	*		*	*		*	*	*	*		*	*	*	*		*	*	*
1	C	A	T	T	C	G	A	A	G	C	G	A	T	C	A	G	G	C	T	A	T	A
2	C	A	T	T	T	G	A	G	A	C	G	A	T	C	A	G	G	C	T	A	T	A
3	C	G	T	T	T	G	A	G	A	C	G	A	T	T	A	G	G	C	C	A	T	A
4	C	A	T	T	C	G	A	G	A	C	G	A	T	C	A	G	G	C	T	A	T	A
outgroup	T	A	C	C	C	A	G	G	A	G	A	T	A	C	G	C	A	T	T	T	A	T

- = non-coding
- = synonymous
- = nonsynonymous

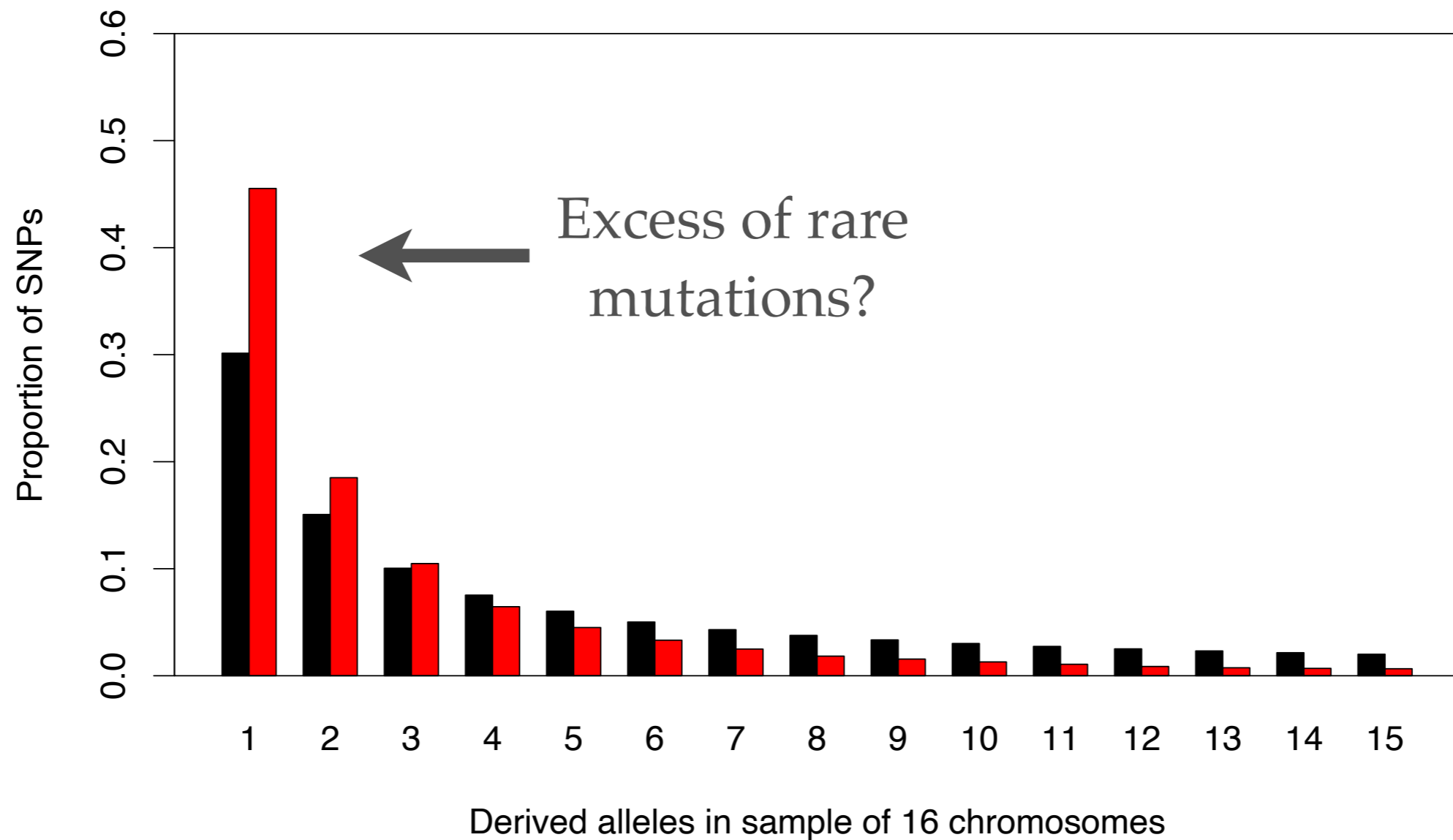
* - Substitution between species

Site-Frequency Spectrum

The proportion of derived mutations at each frequency in a sample of chromosomes



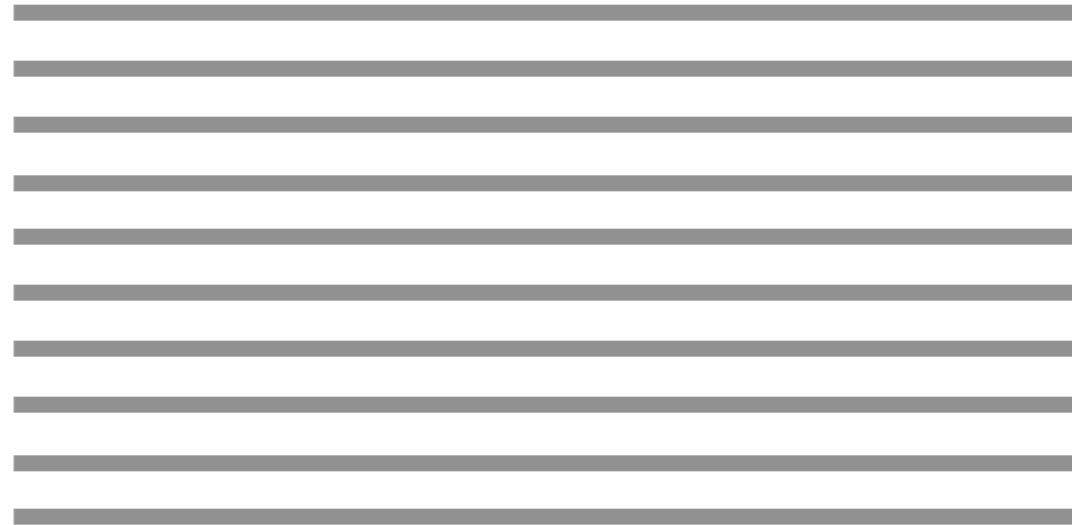
SITE-FREQUENCY SPECTRUM



What evolutionary forces could cause an excess of rare variants? (one-word-answers-please)

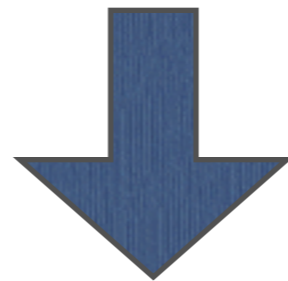
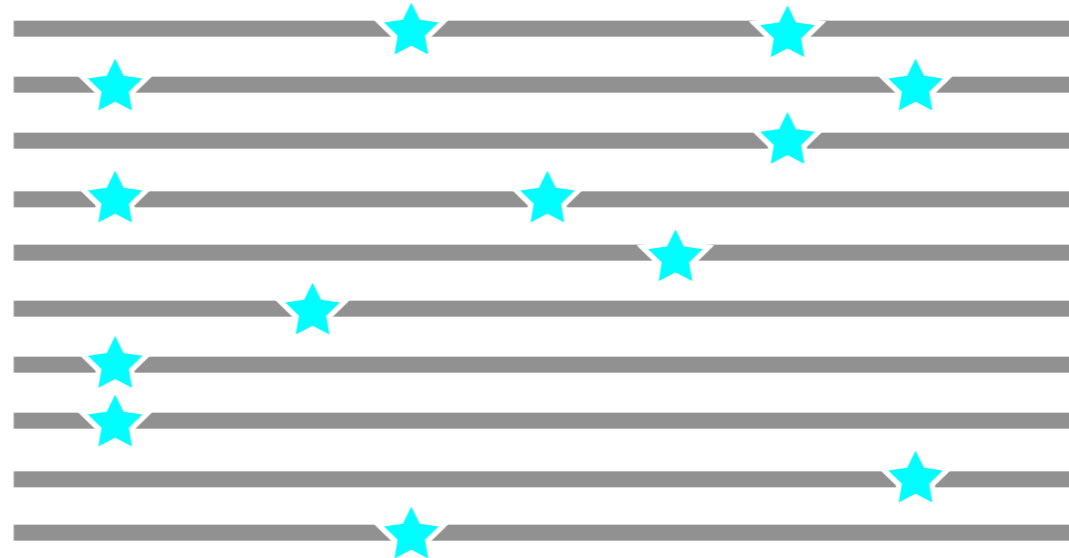
The Effect of Negative Selection

Chromosomes in
a population

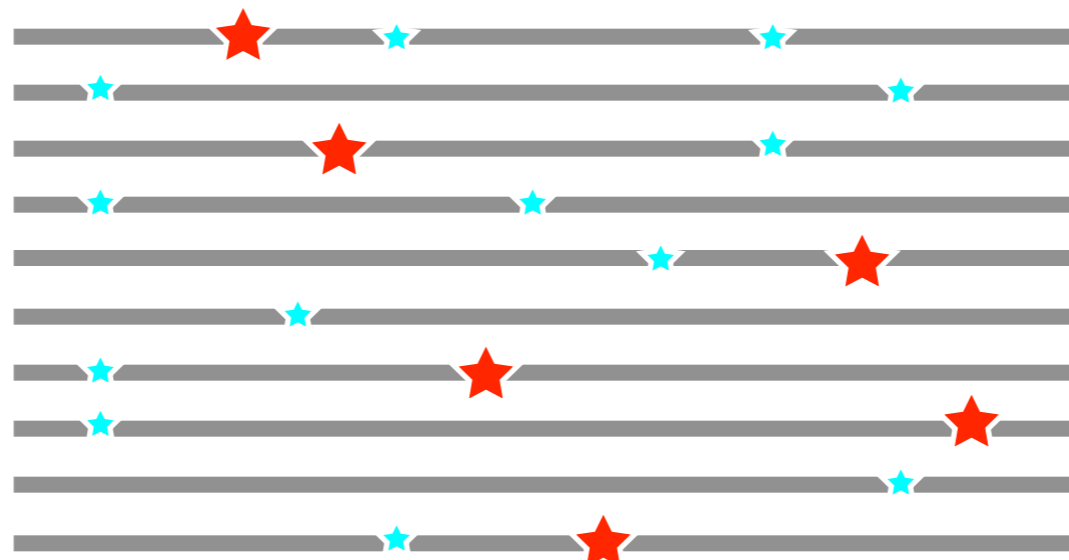


The Effect of Negative Selection

Chromosomes in
a population with
standing variation

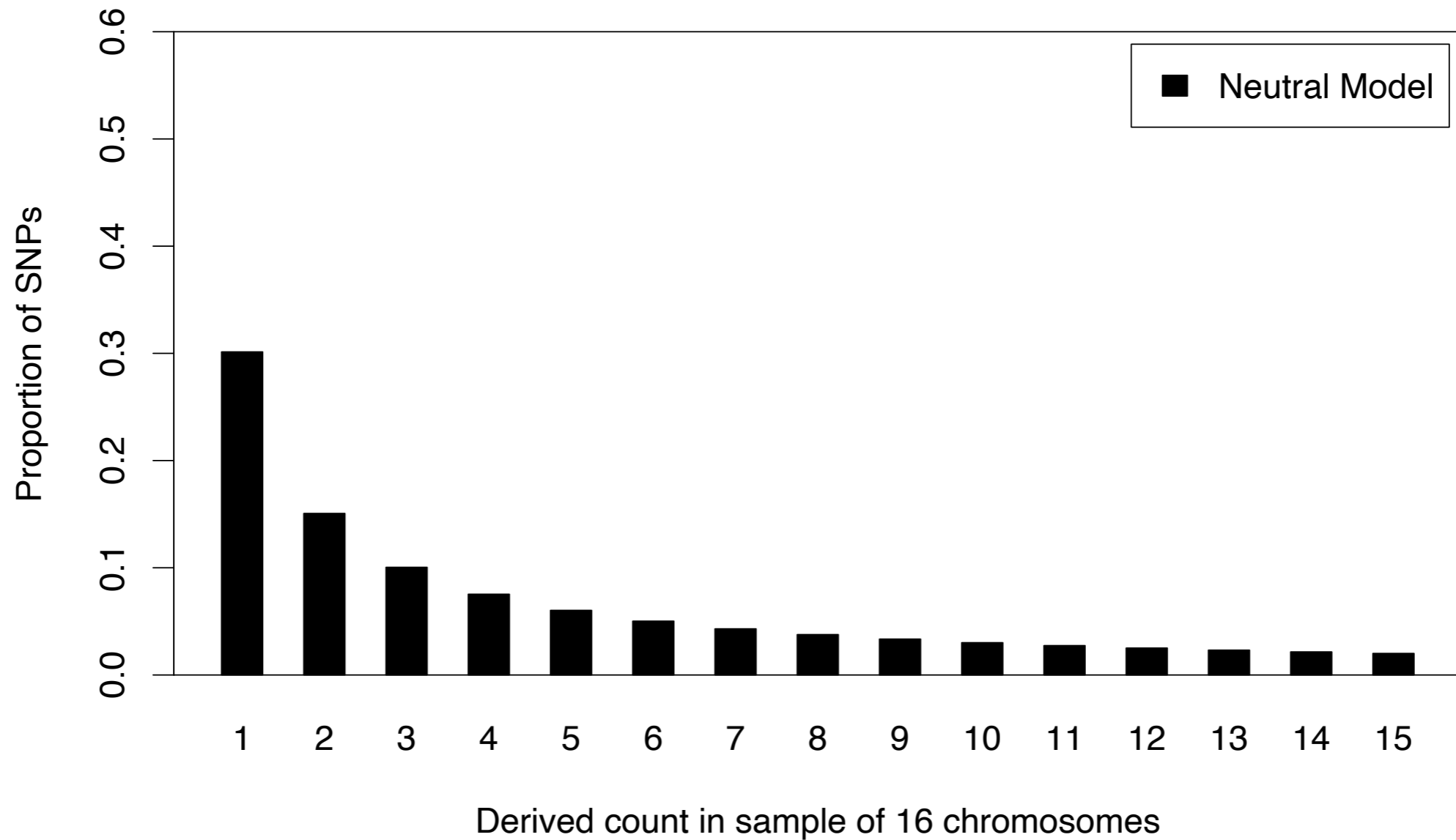


Deleterious
mutations will
arise in the next
generation

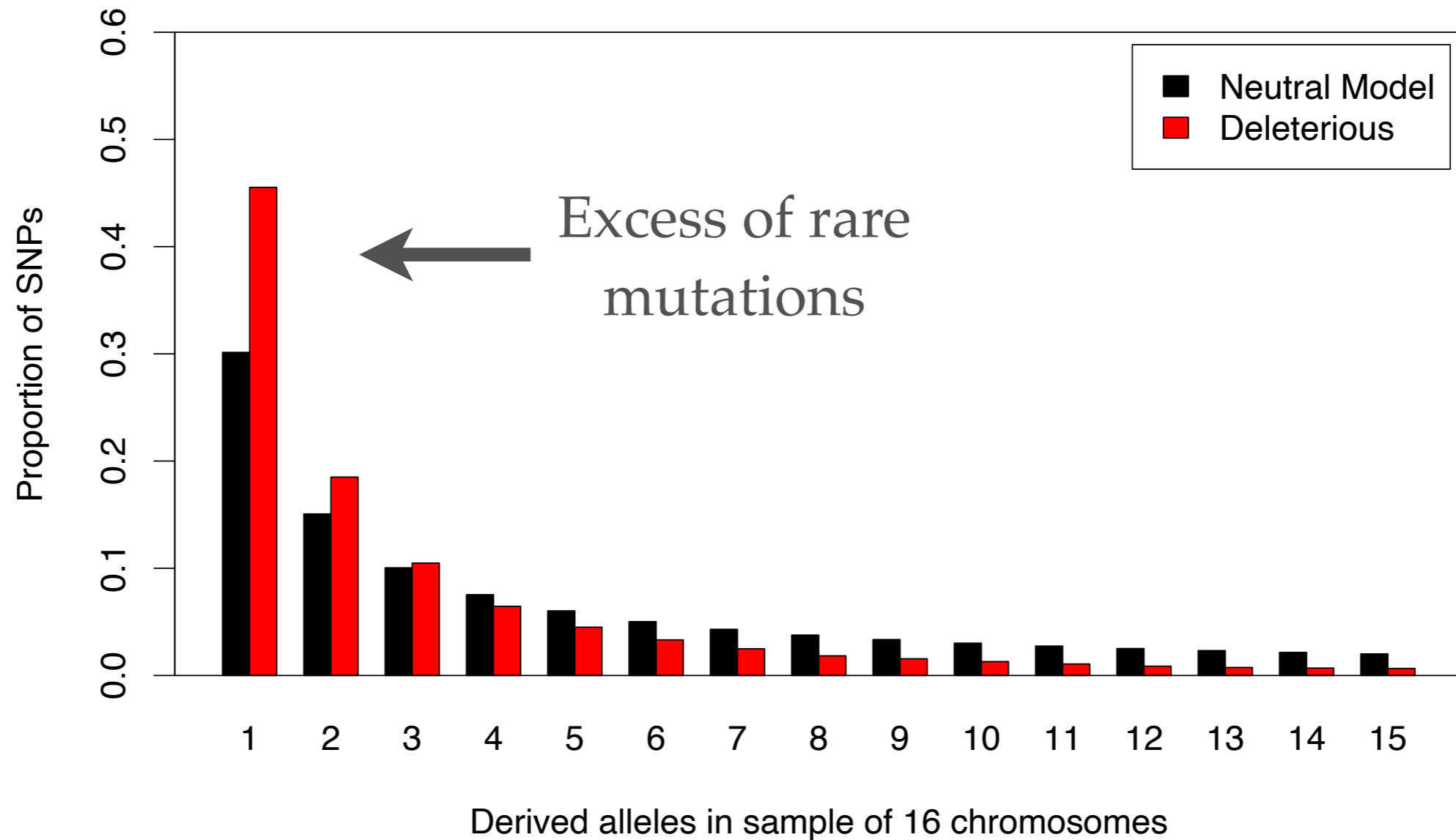


Negative selection:
the action of
natural selection
purging deleterious
mutations.

Site-Frequency Spectrum

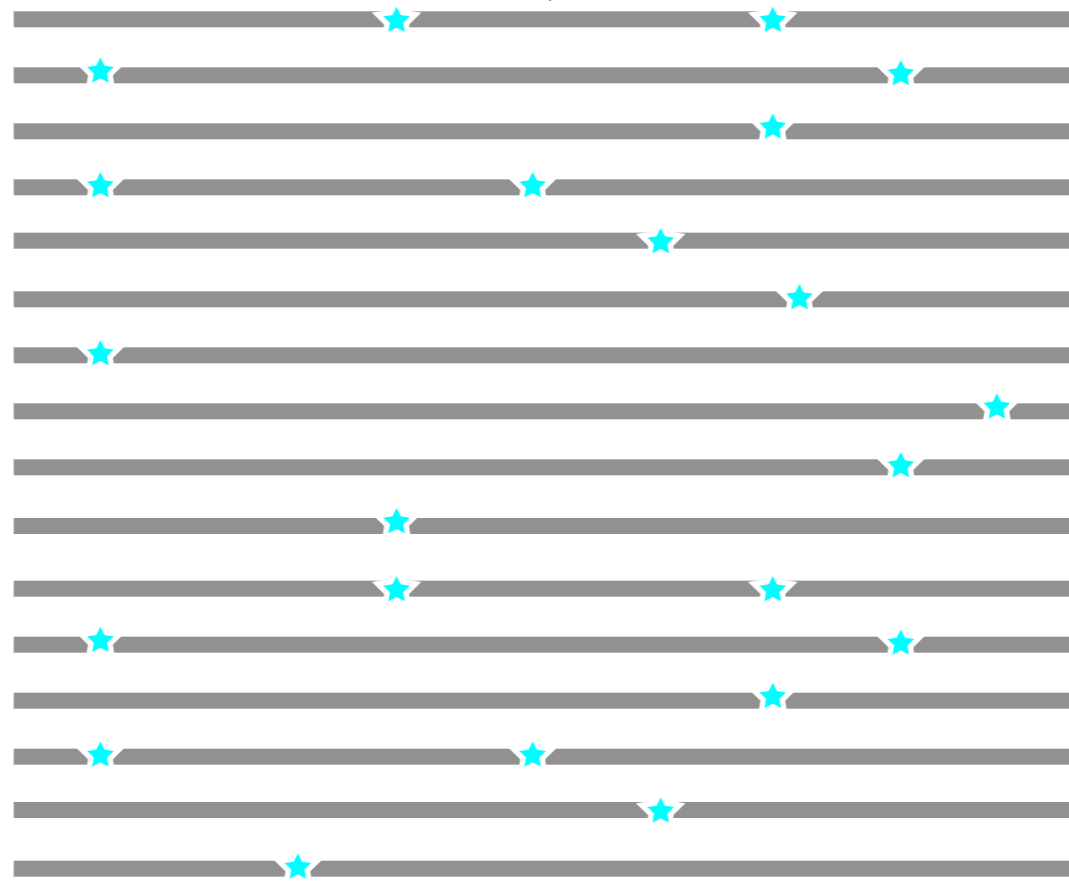


SITE-FREQUENCY SPECTRUM



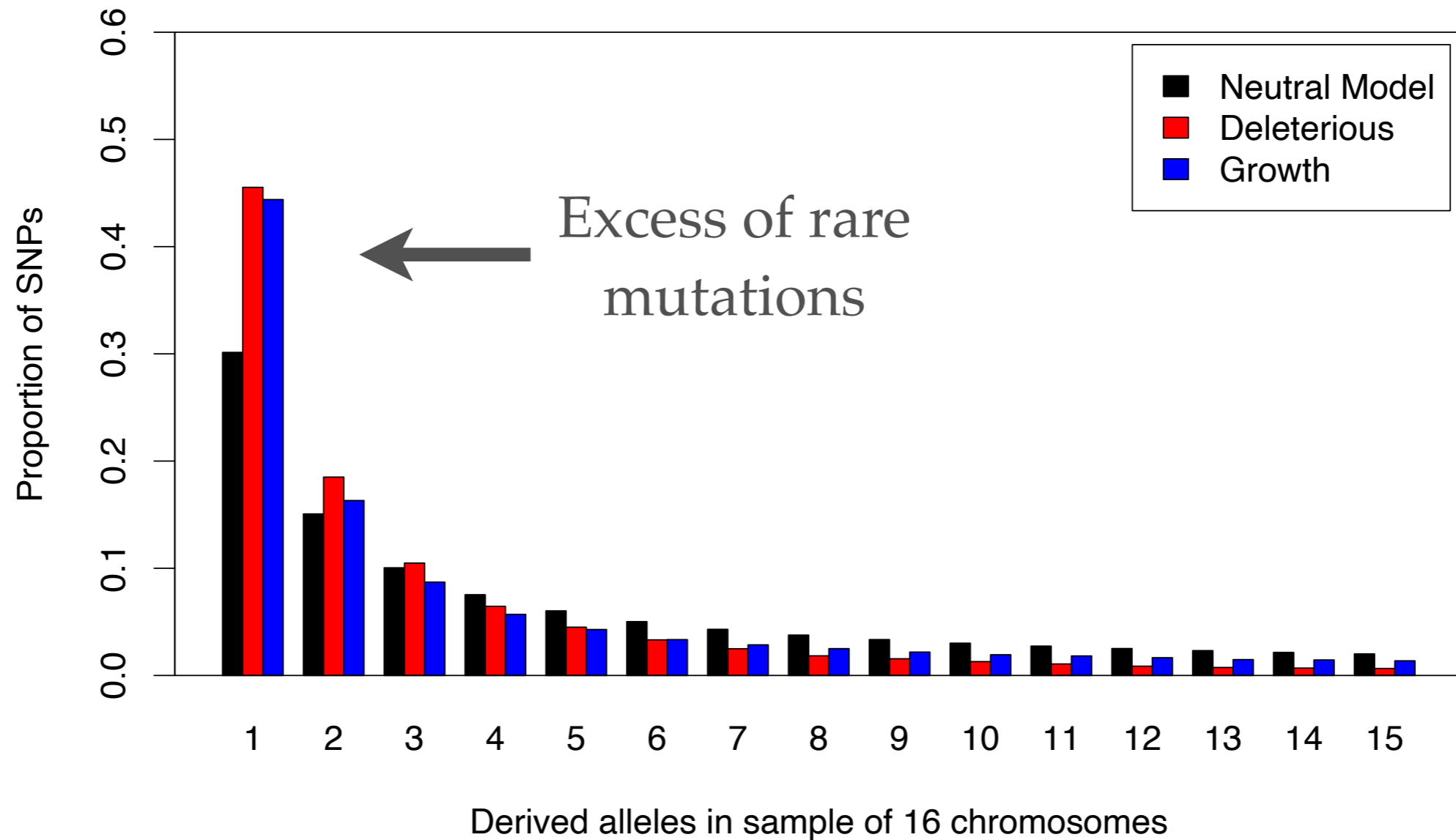
The Effect of Population Growth

Chromosomes in
a population with
standing variation

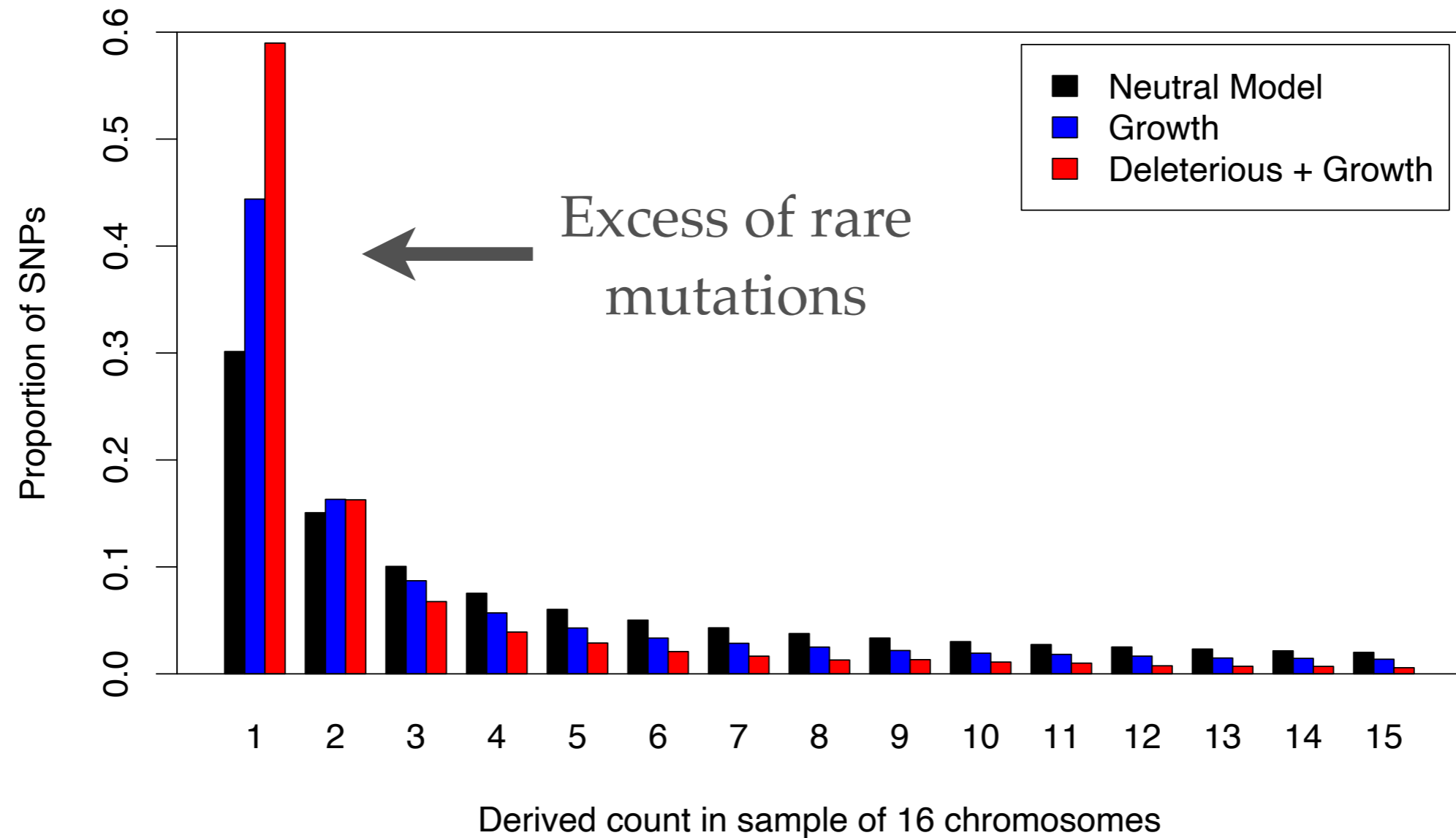


Population Growth:
introduction of new
(rare) variants.

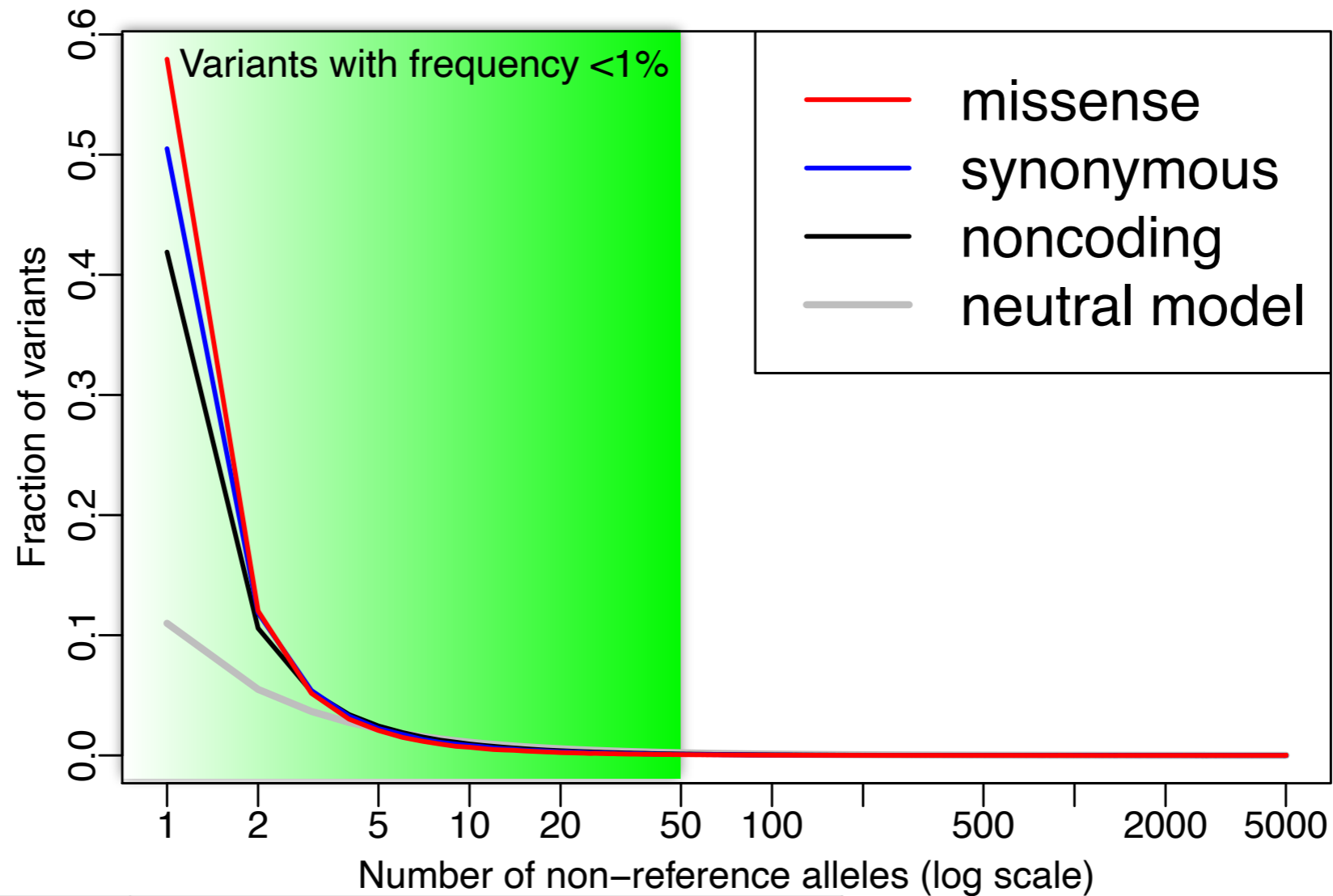
SITE-FREQUENCY SPECTRUM



SITE-FREQUENCY SPECTRUM

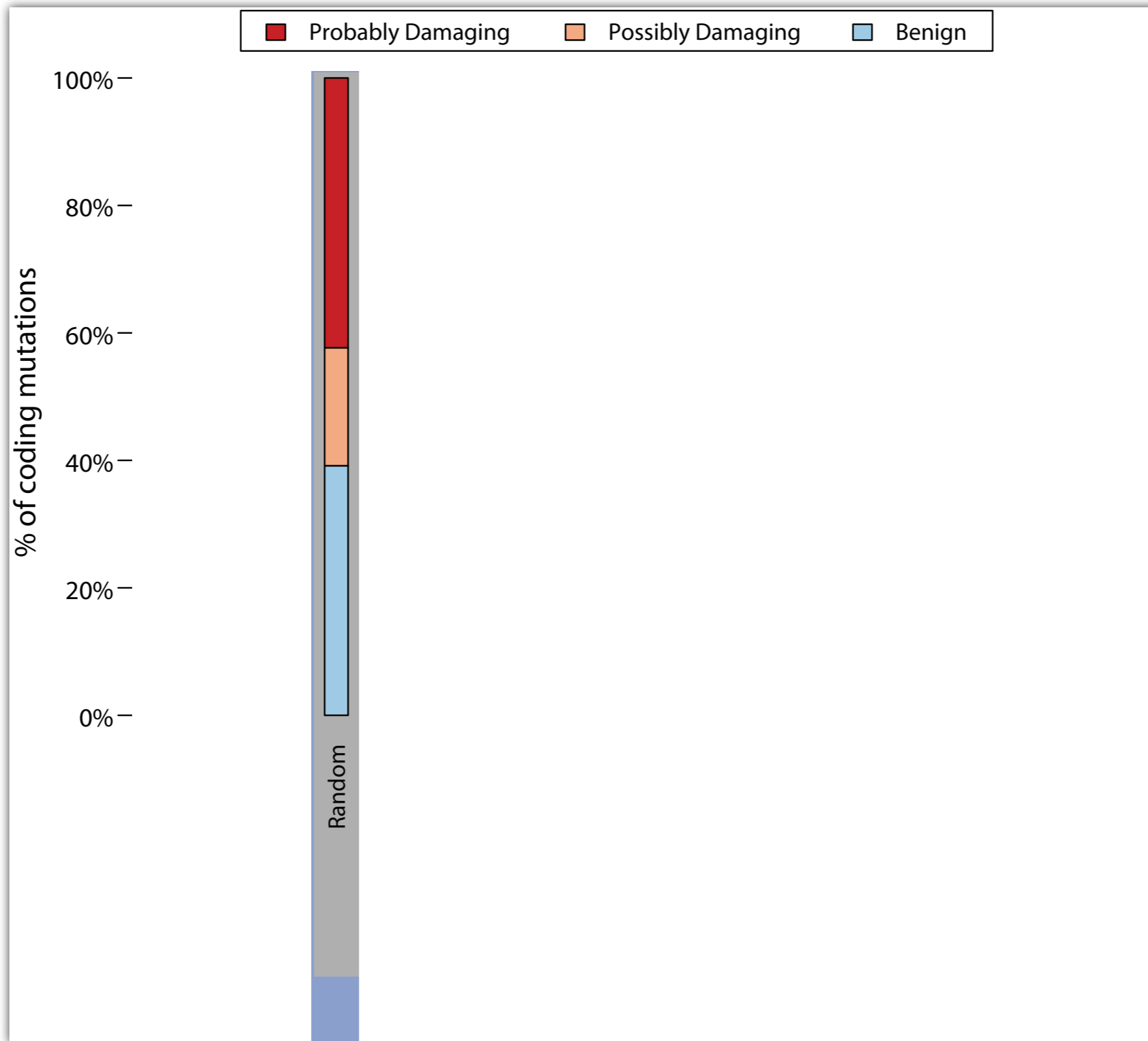


Majority of human genetic variation is rare

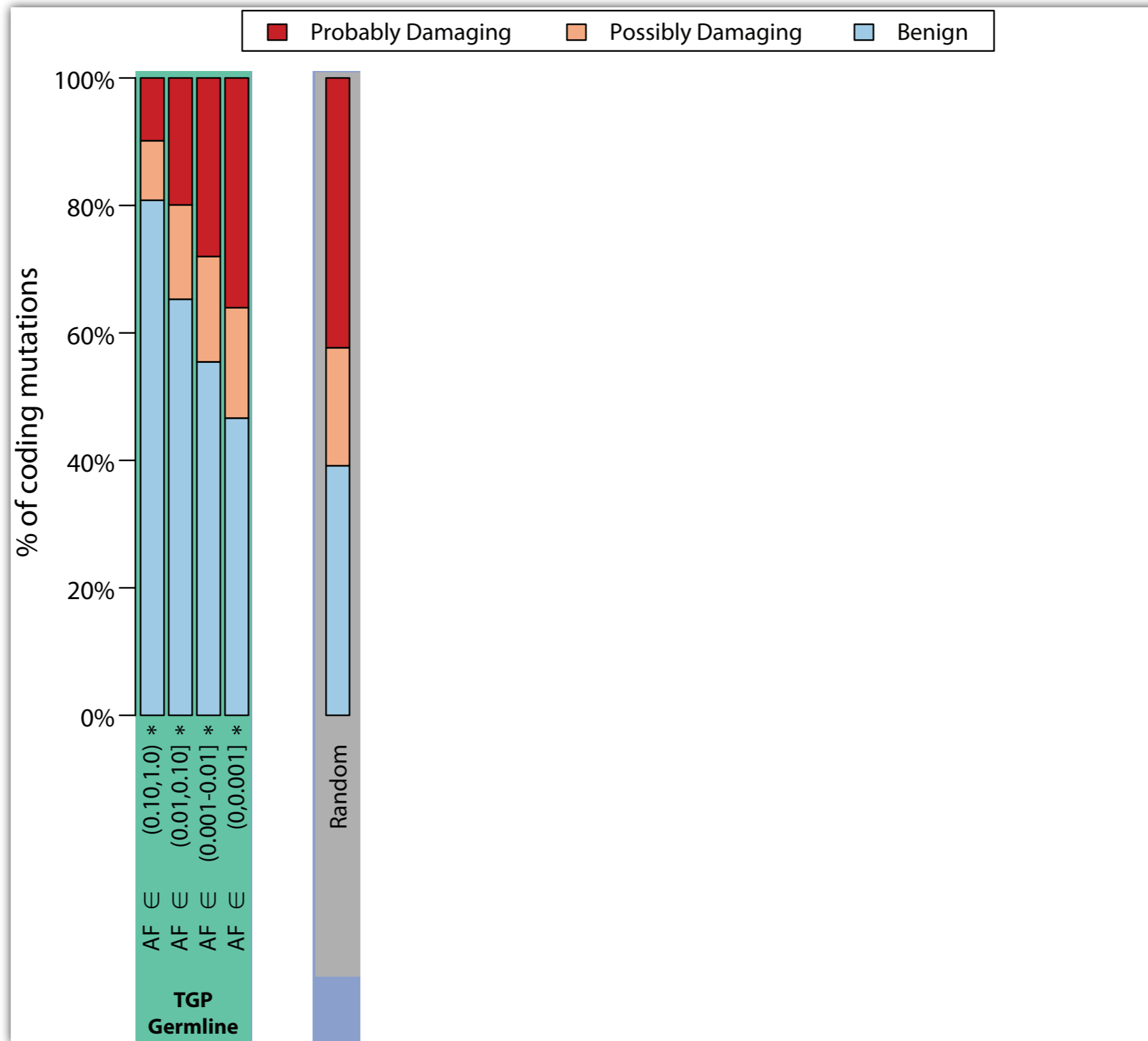


Class	Fraction of variants < 1%
Missense	92.6%
Synonymous	88.5%
Non-coding	82.3%

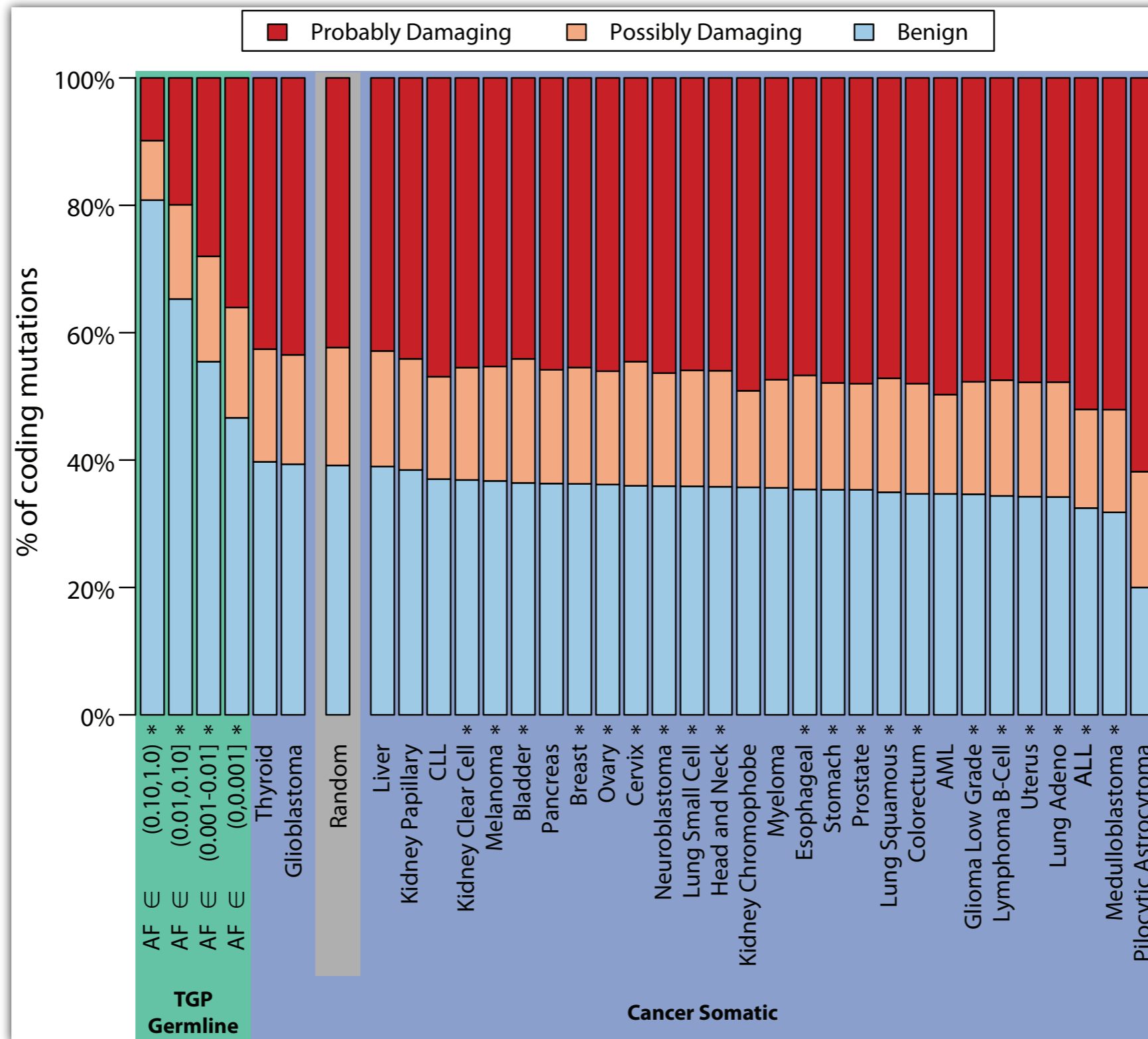
Observed Effect of Selection



Observed Effect of Selection

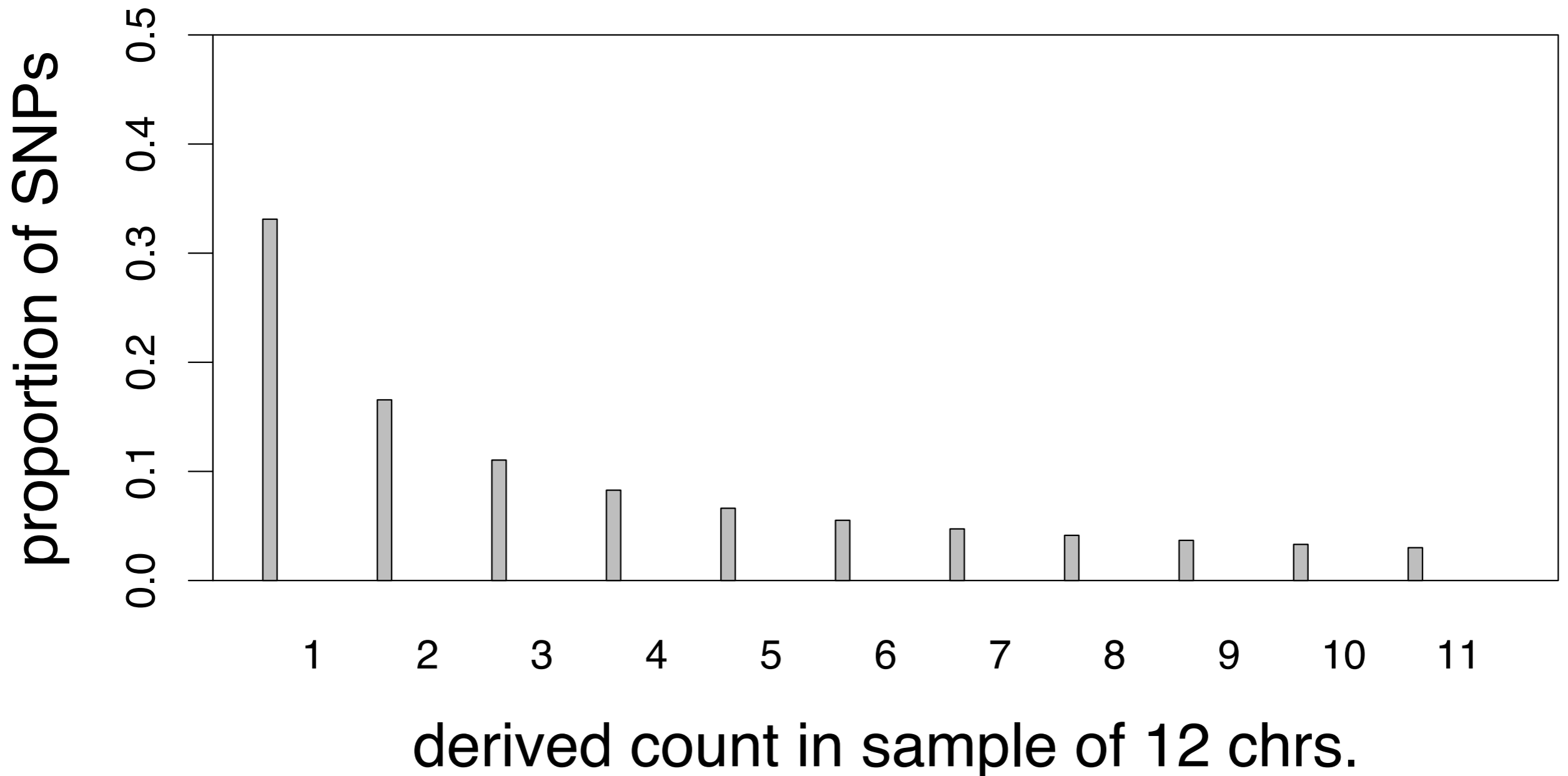


Observed Effect of Selection

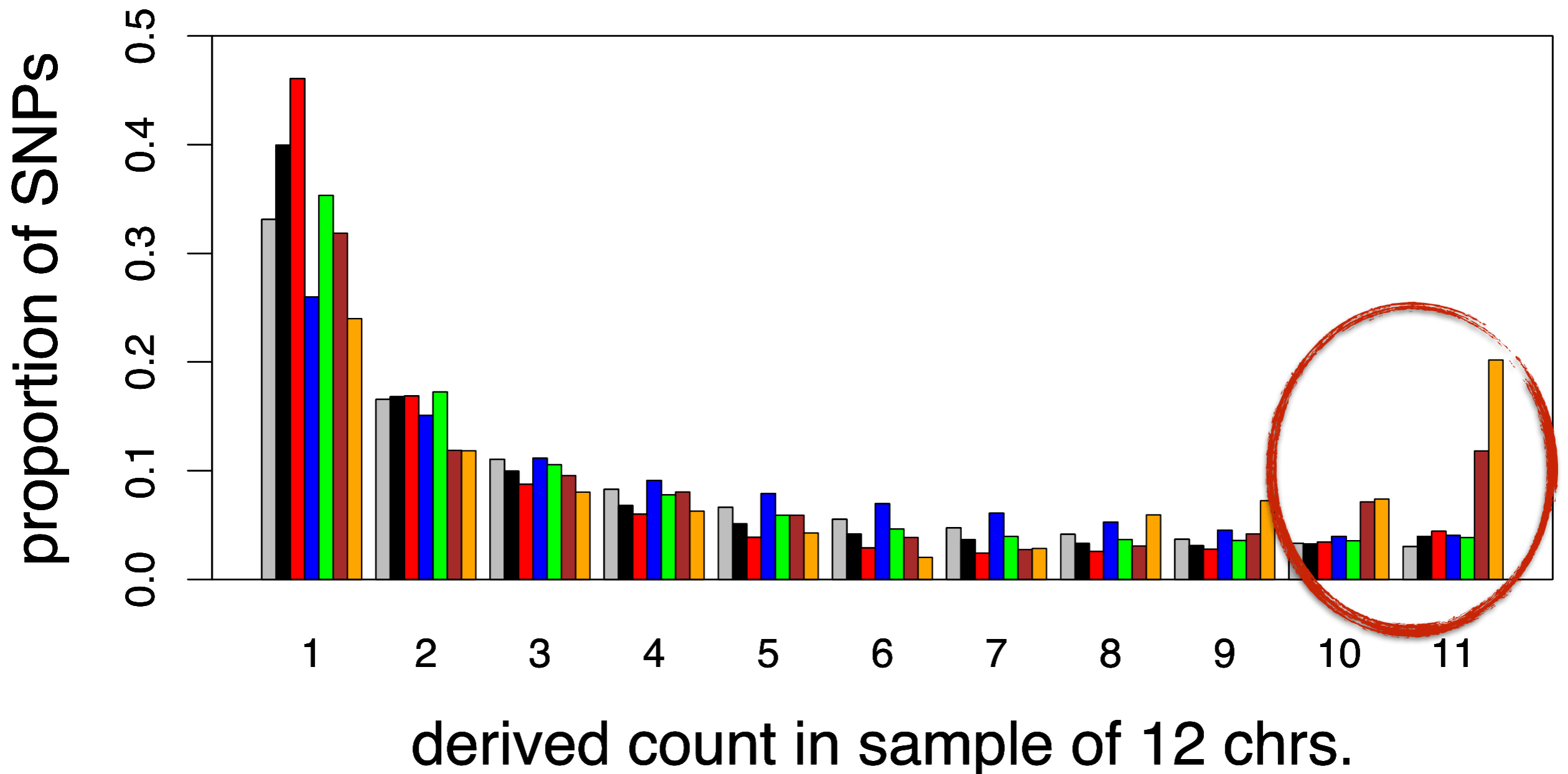
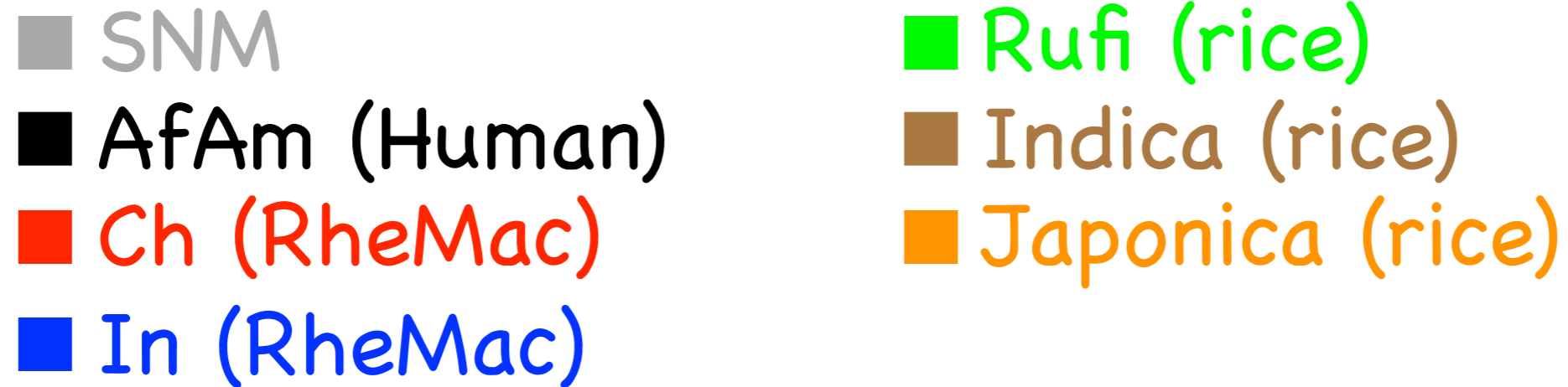


Site-Frequency Spectrum

- The proportion of SNPs at each frequency in a sample of chromosomes.



Site-Frequency Spectrum



**What evolutionary forces could cause an excess of high frequency derived variants?
(one-word-answers-please)**

Site-Frequency Spectrum

■ SNM

■ AfAm (Human)

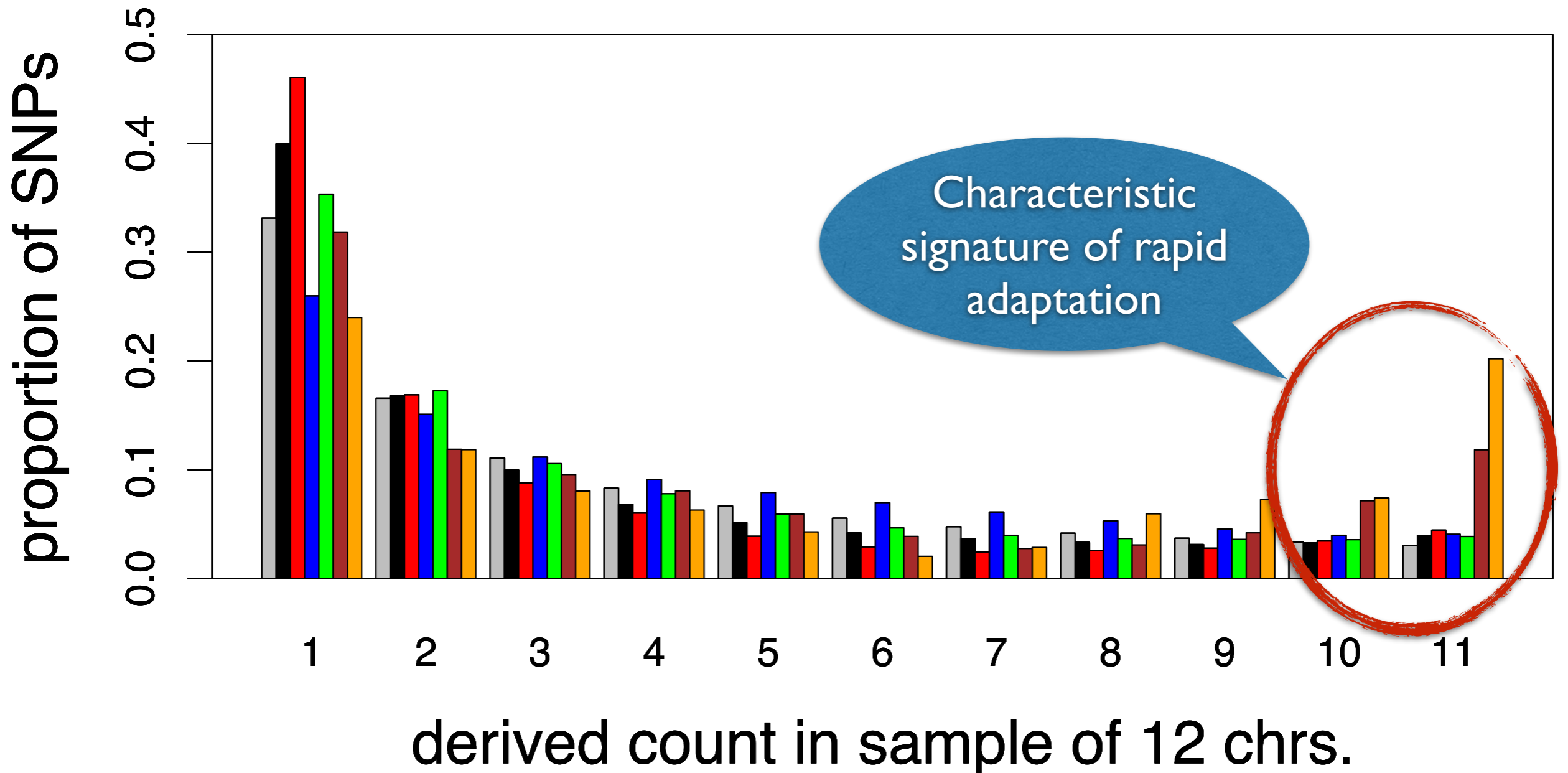
■ Ch (RheMac)

■ In (RheMac)

■ Rufi (rice)

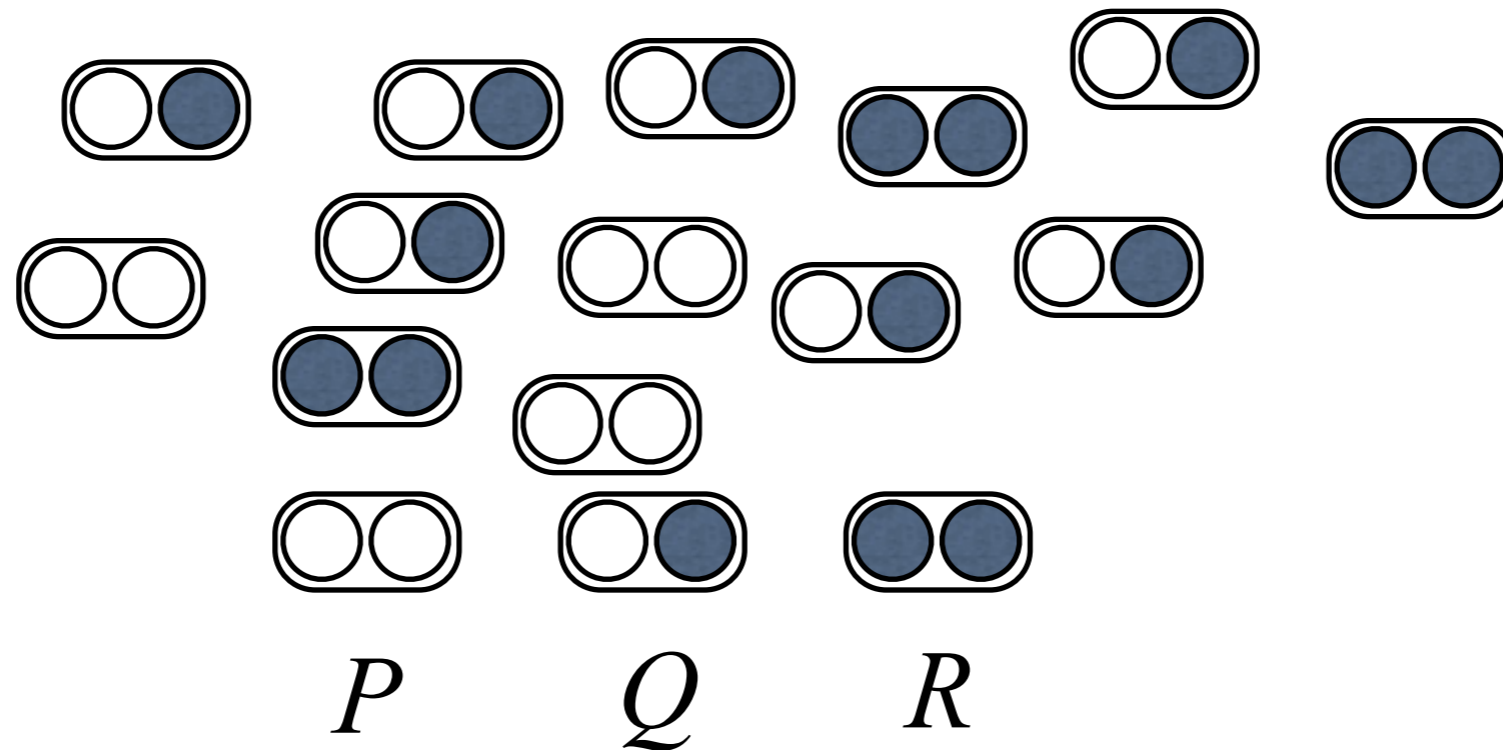
■ Indica (rice)

■ Japonica (rice)



Population Genetics

- Imagine a population of diploid individuals



- Principles of **random mating**:
 - Any two individuals are equally likely to mate and reproduce to populate the next generation.
 - Either chromosome is equally likely to be passed on.

Hardy-Weinberg Principle



Godfrey H. Hardy:
1877-1947



Wilhelm Weinberg:
1862-1937

What are the assumptions of the Hardy-Weinberg Principle?

Hardy-Weinberg Principle



Godfrey H. Hardy:
1877-1947



Wilhelm Weinberg:
1862-1937

● Assumptions:

- Diploid organism
- Sexual reproduction
- Non-overlapping generations
- Only two alleles
- Random mating
- Identical frequencies in males/females
- Infinite population size
- No migration
- No mutation
- No natural selection

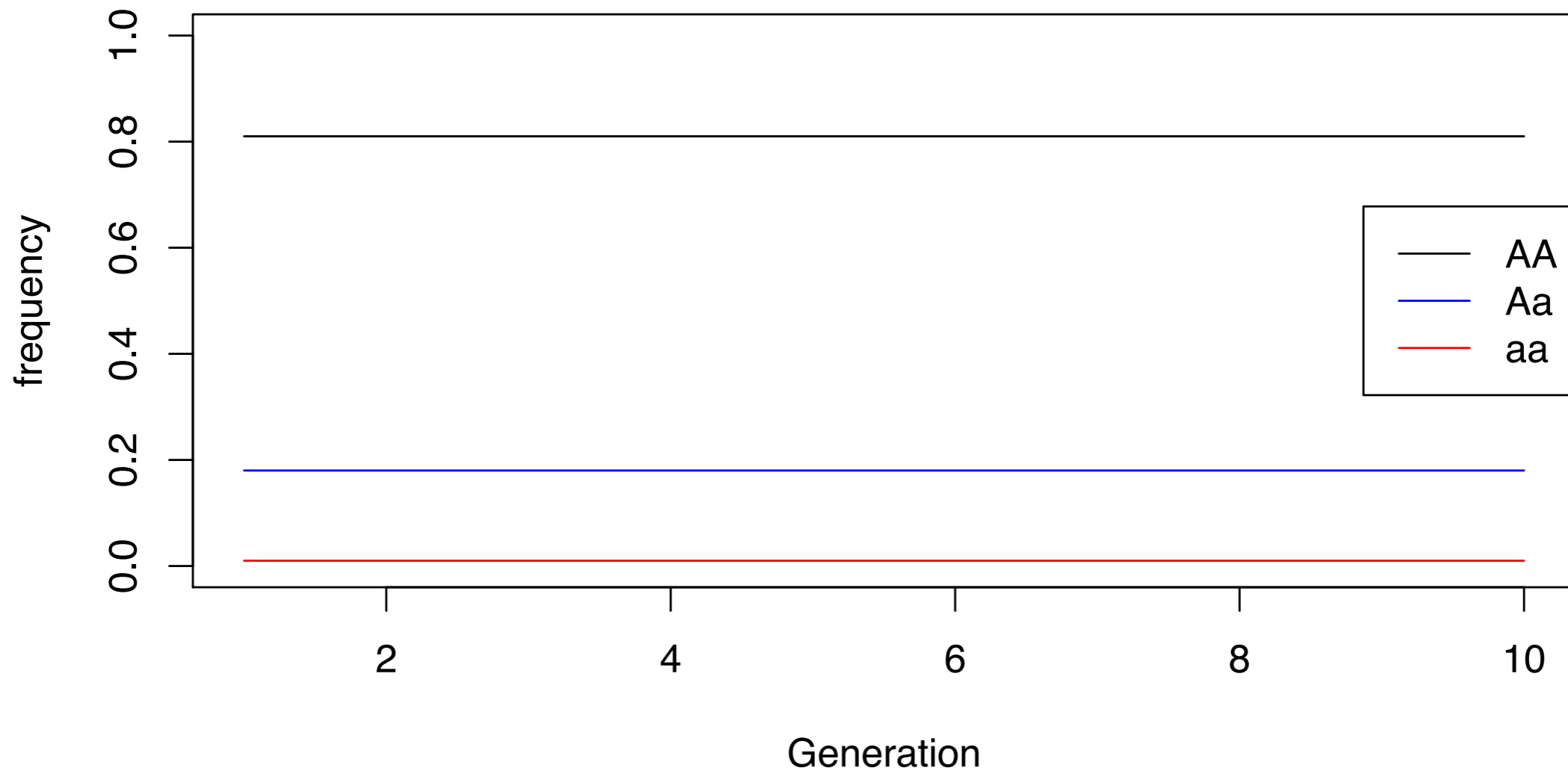
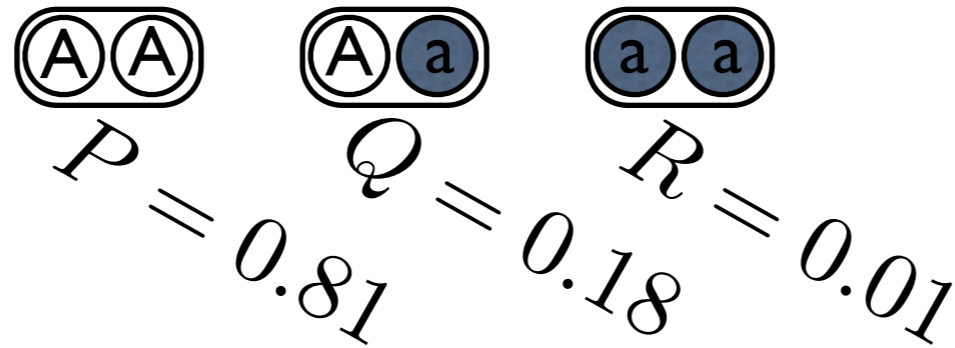
● Conclusion I:

Both allele AND genotype frequencies will remain constant at **HWE** generation after generation... forever!

$$P=p^2$$
$$Q=2p(1-p)$$
$$R=(1-p)^2$$

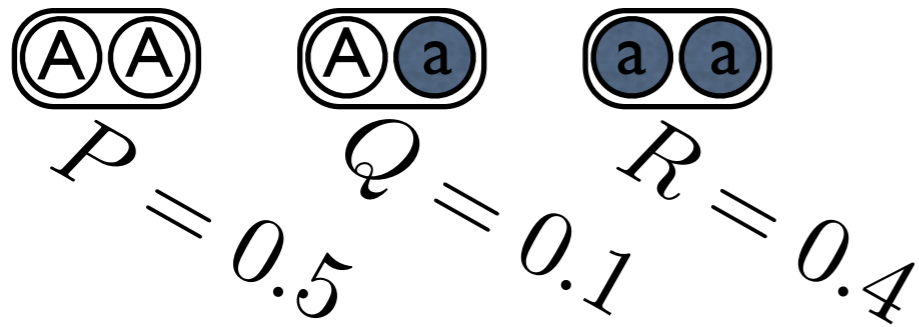
Hardy-Weinberg Principle

- Imagine a population of diploid individuals



Hardy-Weinberg Principle

- Imagine a population of diploid individuals

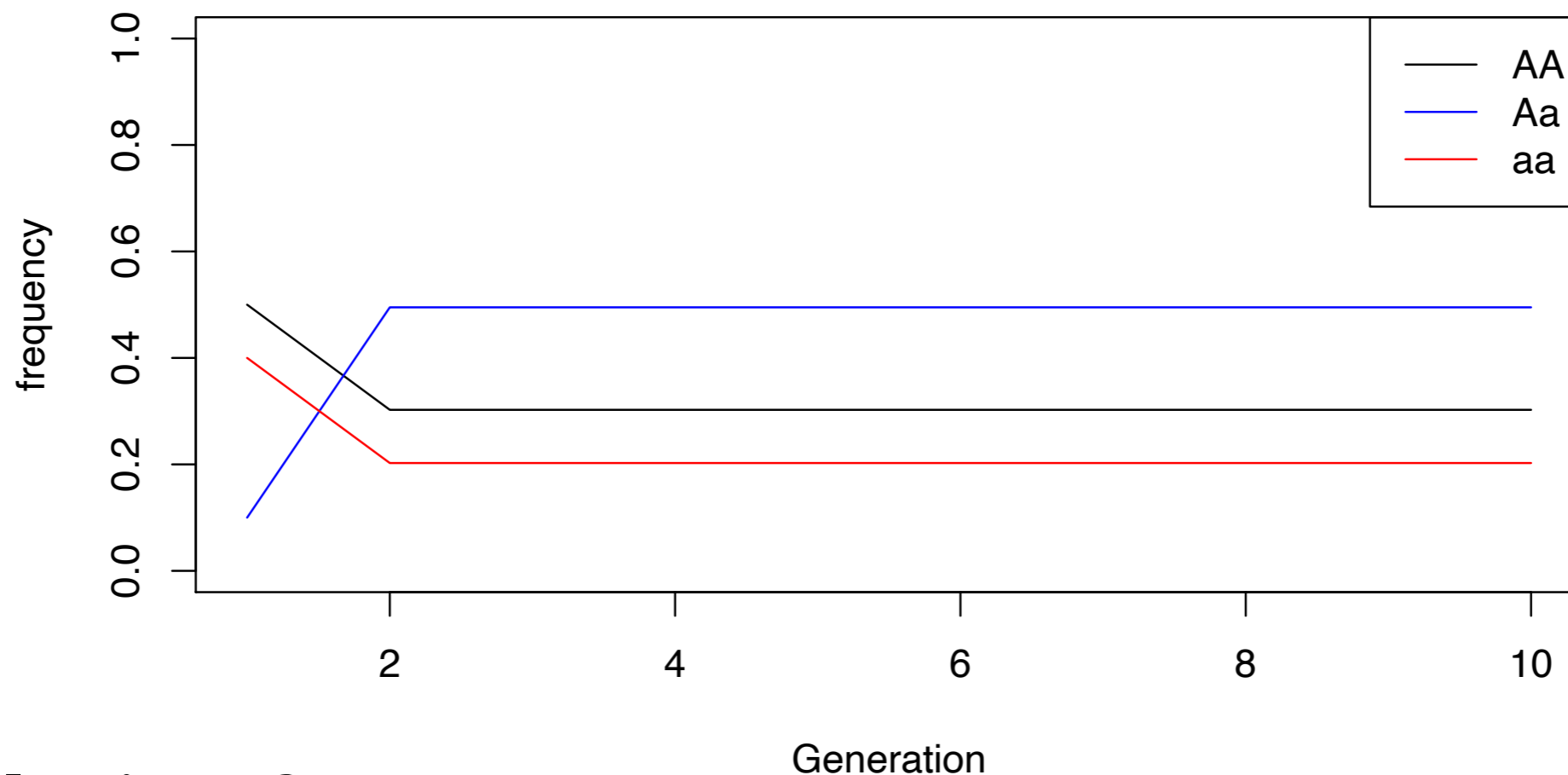


$$p = P + Q/2 = 0.55$$

$$p^2 = 0.3025$$

$$2p(1 - p) = 0.495$$

$$(1 - p)^2 = 0.2025$$



- Conclusion 2:** A single round of random mating will return the population to HWE frequencies!

Hardy-Weinberg Principle



Godfrey H. Hardy:
1877-1947

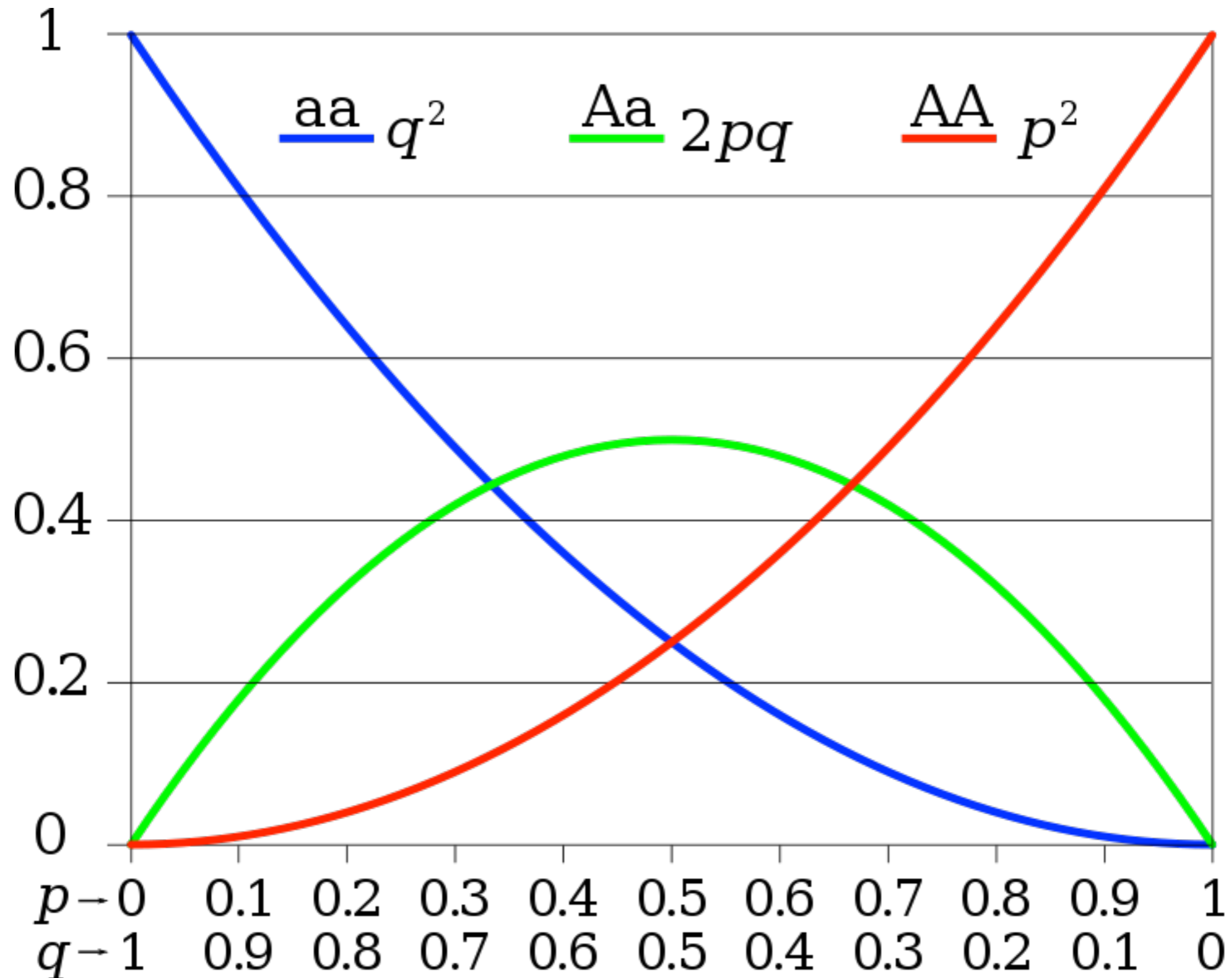


Wilhelm Weinberg:
1862-1937

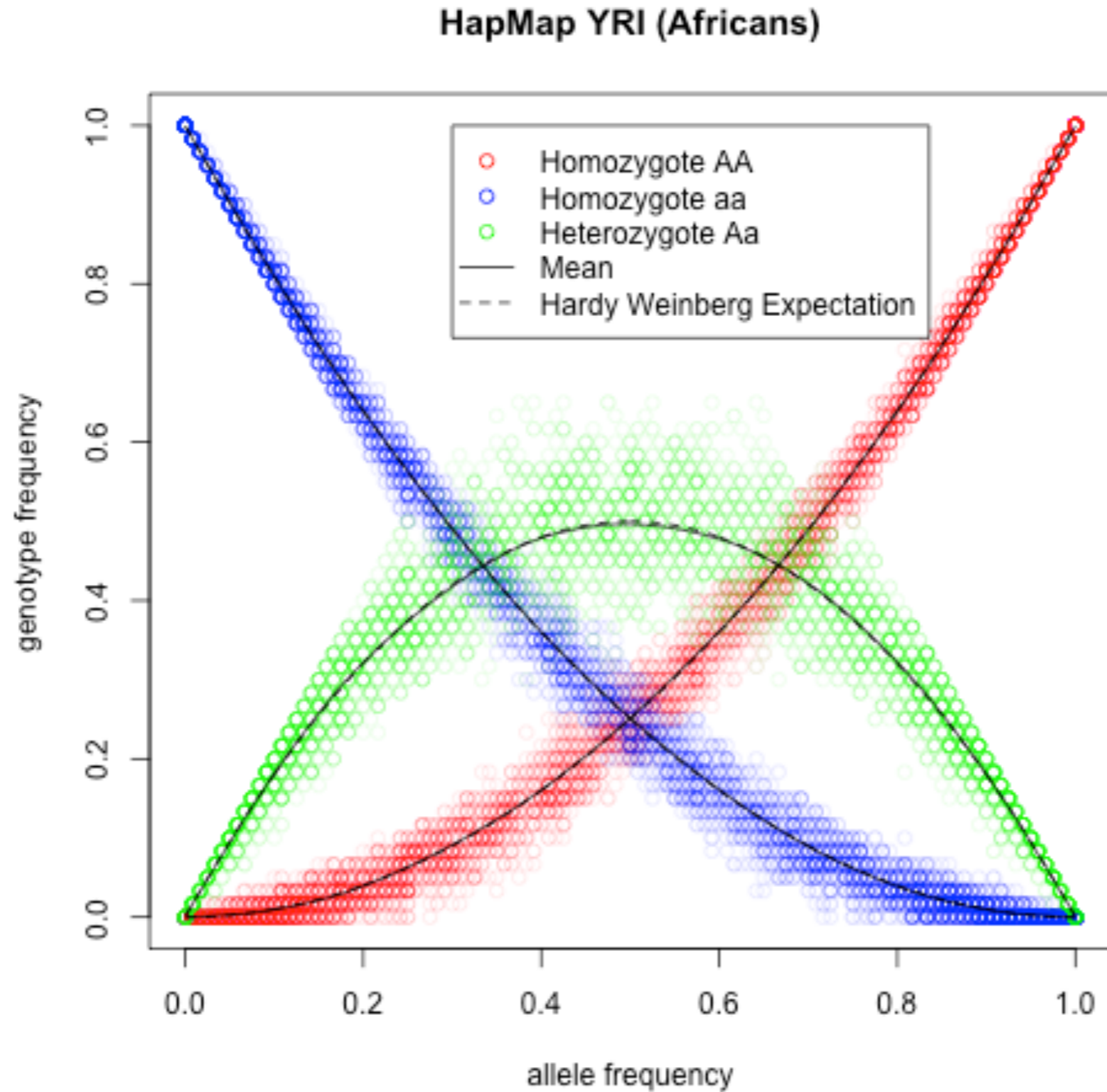
● Assumptions:

- Diploid organism
- Sexual reproduction
- Non-overlapping generations
- Only two alleles
- Random mating
- Identical frequencies in males/females
- Infinite population size
- No migration
- No mutation
- No natural selection

Hardy-Weinberg Equilibrium



Hardy-Weinberg Equilibrium



Summary

- Hardy-Weinberg Equilibrium requires many assumptions, all of which are routinely violated in natural populations.
- Nevertheless, the vast majority of variants are in HWE.
 - Deviations almost always due to technical artifacts!
- Natural selection changes the expected allele frequency in the next generation.
 - But drift still acts in finite populations!