

Population genetic inference

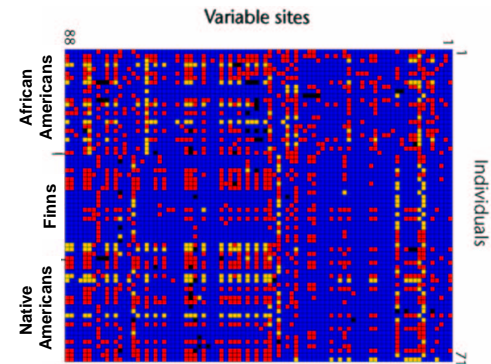
Date	Topic	
22 nd Jan	Good questions in population genetics	GM
29 th Jan	Principles of population genetic inference	GM
5 th Feb	Recombination in the coalescent	JH
12 th Feb	Natural selection	GM
19 th Feb	Demographic models	GM
26 th Feb	Combinatorics of the coalescent	JH
5 th March	Population genetics of disease mutations	GM
12 th March	Model organisms	GM

Books

- Crow JF & Kimura M. 1970. **An introduction to population genetics theory**. Harper and Row, New York.
- Gillespie JH. 1998. **Populations genetics: a concise guide**. The Johns Hopkins University Press, Baltimore.
- Hartl DL & Clark AG (1989). **Principles of population genetics**. Sinauer Associates, Sunderland, Mass.
- Li, W-H. (1997) **Molecular Evolution**. Sinauer Associates, Sunderland, Mass.

1

Good questions in population genetics

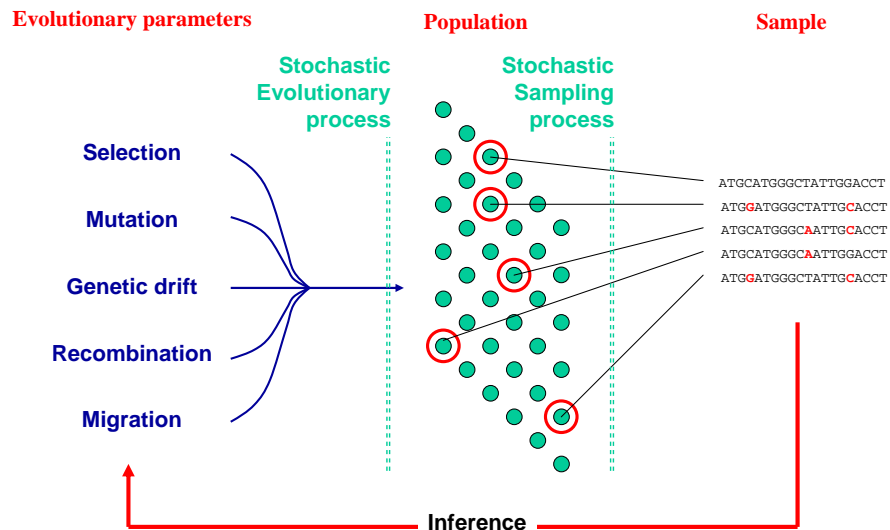


DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene
Nickerson, *et al.* 1998 *Nature Genetics* 19, 233 - 240

- Is there an association between DNA sequence variation and the disease phenotype?
- What do the sequences tell us about human history?
- How has natural selection shaped diversity in the gene?

2

Population genetic inference



3

Historical obsessions about variation

- How is variation inherited?
- Can particulate inheritance explain quantitative traits?
- What maintains variation in populations?
- What is the molecular basis of variation?
- Why is there so much variation?

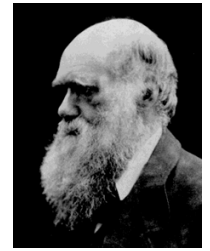
4

The early history of population genetics

Date	Event
1859	Darwin's <i>Origin of Species</i>
1856-63	Mendel's experiments on peas
1900	Rediscovery of Mendel's laws
1909	Nilsson-Ehle's experiments on wheat
1912-1920	Pearl, Jennings and Wright's work on inbreeding
1915	Morgan's experiments on <i>Drosophila</i>
1918	Fisher's paper on phenotypic correlations between relatives
1918	Sturtevant's artificial selection experiments on <i>Drosophila</i>
1930	Fisher's <i>The Genetical Theory of Natural Selection</i> (Fundamental theorem)
1931	Wright's <i>Evolution in Mendelian populations</i>
1932	Haldane's <i>The Causes of Evolution</i>
1955	Kimura diffusion equation solution to the distribution of allele frequencies

5

What Darwin said



Organisms produce too many offspring

Heritable differences exist in traits influencing the adaptation of an organism to its environment

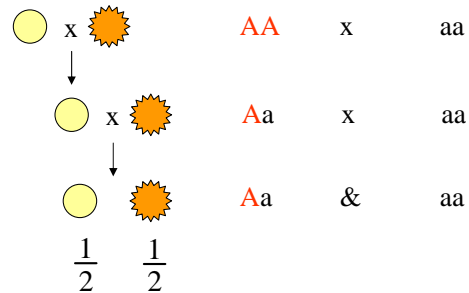
Organisms that are better adapted have a higher chance of survival



Q: How can variation be inherited?

6










Mendel's peas



Q: How can particulate inheritance explain quantitative traits?

7

Nilsson Ehle's wheat

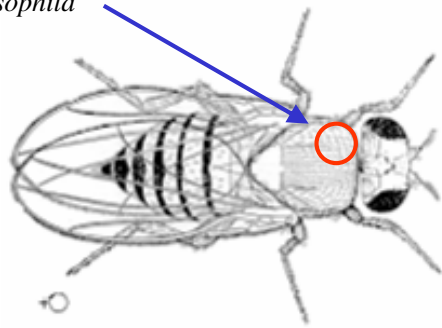
Genotype	AA	Aa	aa
BB			
Bb			
bb			

8

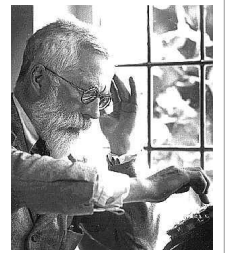
Morgan and his flies

(+ Sturtevant, Muller and Bridges)

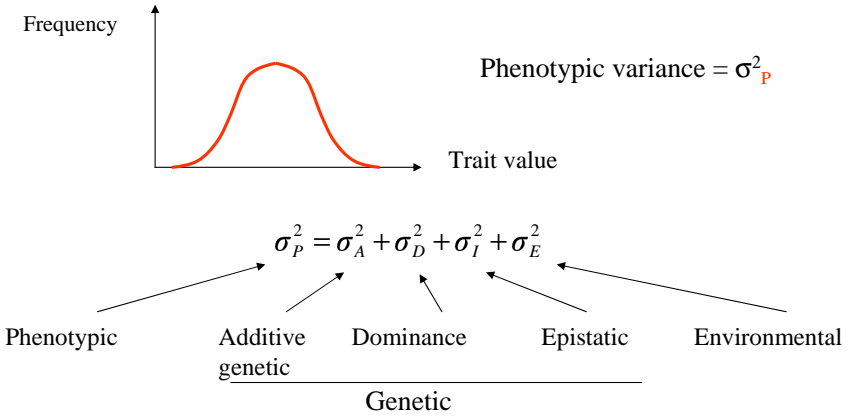
- Discovered crossing-over (cM)
- Proved chromosomes carried hereditary factors
- Showed heritability of bristle number in *Drosophila*



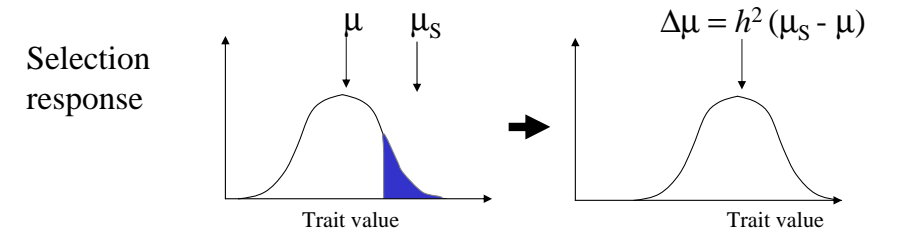
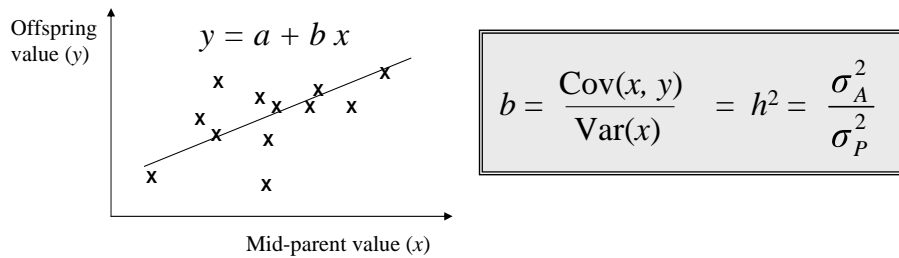
Fisher's results on genetic variation



- Three types of quantitative trait
 - Continuous (weight, height, milk yield)
 - Meristic (bristle number in *Drosophila*)
 - Discrete with continuous liability (disease susceptibility)

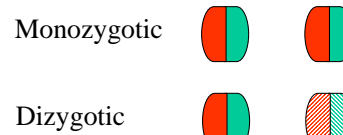


Estimating the genetic component of quantitative traits

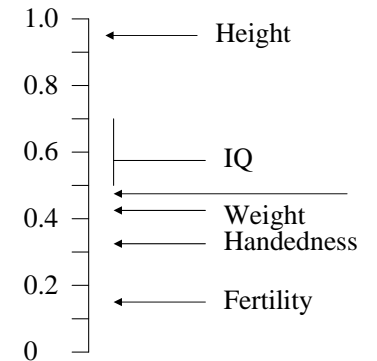


Heritabilities of human traits

Twin studies



$$h^2 = \frac{r_M - r_D}{1 - r_D}$$



Twin concordance in human disease

Disease	Concordance		Genetic Determinism
	DZ	MZ	
Cancer	6.8	2.6	0.23 - 0.33
Arterial hypertension	25.0	6.6	0.53 - 0.62
Manic-depressive psychosis	67.0	5.0	1.04 - 1.05
Tuberculosis	37.2	15.3	0.53 - 0.65

From Cavalli-Sforza & Bodmer (1971)

Q: Why do different traits have different levels of heritability?

13

Selection, mutation and genetic drift

Factors promoting variation	Factors reducing variation
Mutation	Genetic drift (random sampling of gametes)
Natural Selection (balancing selection)	Natural Selection (stabilising selection and adaptive evolution)

Q: When should one force dominate the others?

14

The neo-synthesis

- RA Fisher
 - *The Genetical Theory of Natural Selection* (1930)
 - Fisher's fundamental theory
 - Geometric model of adaptation
 - The concept of likelihood in statistical analysis
 - Experimental design
- JBS Haldane
 - *The Causes of Evolution* (1932)
 - Fixation probabilities of advantageous alleles
 - Theory of sex-linked loci
 - Eloquent exponent of the theory of evolution by natural selection
- Sewall Wright
 - Evolution in Mendelian populations (1931)
 - Developed the use of diffusion theory in population genetics
 - Importance of genetic drift
 - Selection at multiple-loci
 - Shifting-balance theory of evolution
 - Four volume *Evolution and the genetics of populations* (1968-1978)



15

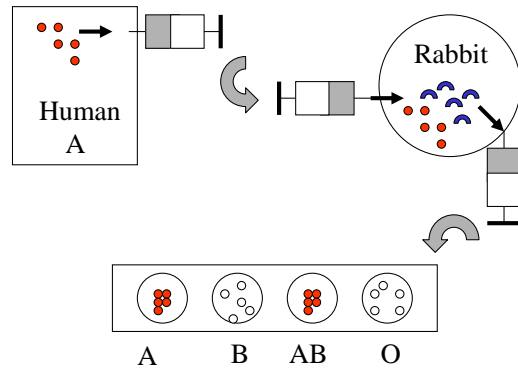
Population genetics before molecular biology

- The (sort of) consensus:
 - Adaptation through natural selection can be achieved even when the selection coefficients are immeasurably small
 - The evolutionary process is highly variable due to the stochastic nature of mutation and reproduction in finite populations
 - Variation is maintained by a balance between natural selection and mutation

Q: What is the molecular basis of variation?

16

Serological techniques for detecting variation



17

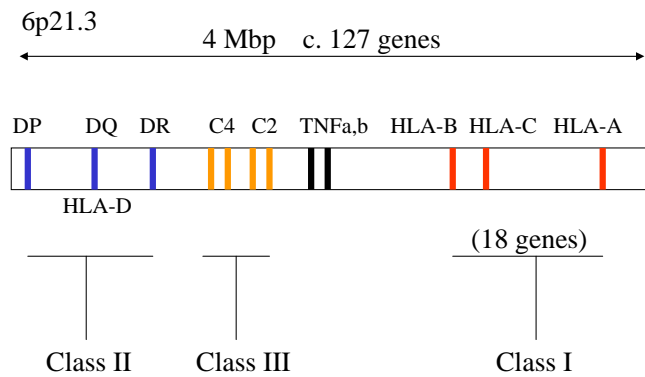
Polymorphic blood groups in the white English population (no. types)

ABO	(4)	Kidd	(3)
Rh	(7)	Dombrock	(2)
MNS	(6)	Auberger	(2)
P	(3)	Xg	(2)
Secretor	(2)	Sd	(2)
Duffy	(3)	Lewis	(2)

Pr{2 people same blood type} \approx 3 in 10,000

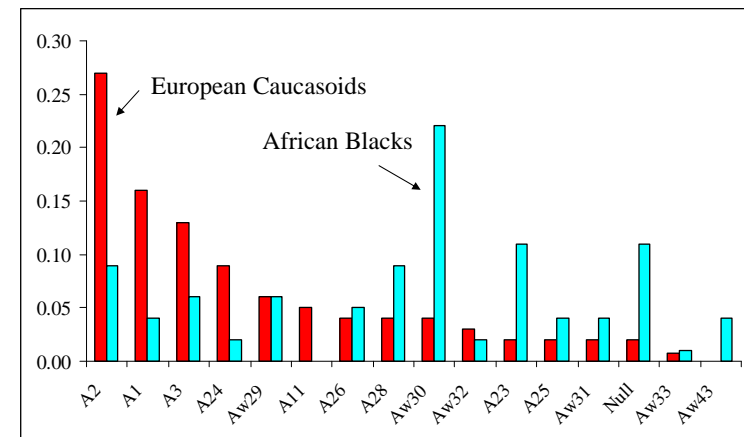
18

HLA diversity at the MHC locus



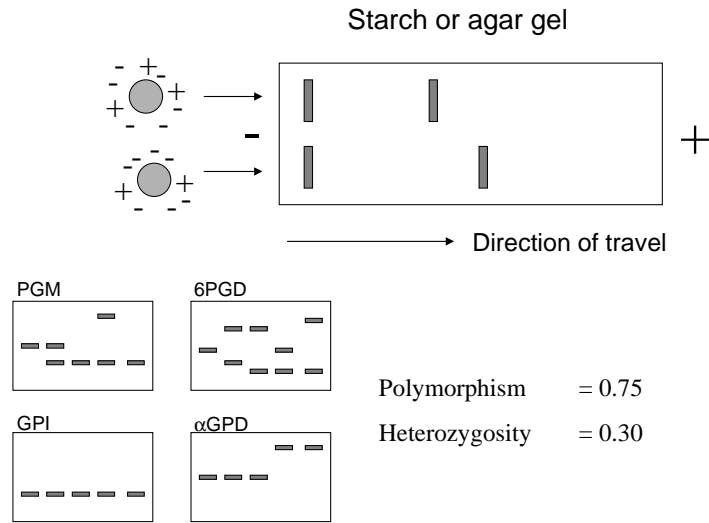
19

HLA-A



20

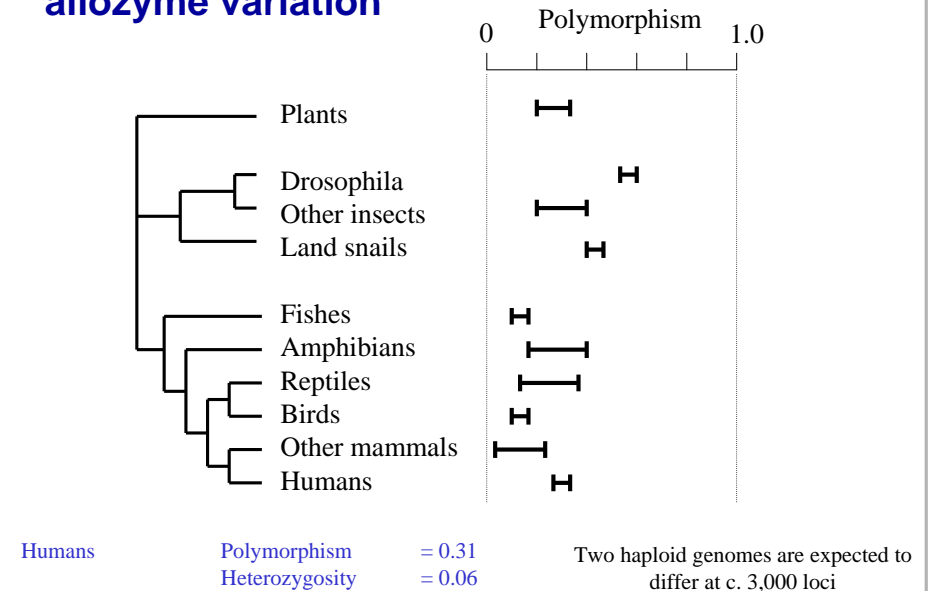
Protein electrophoresis



Lewontin and Hubby (1966)
Harris (1966)

21

The phylogenetic distribution of allozyme variation



22

The rise of the neutral theory

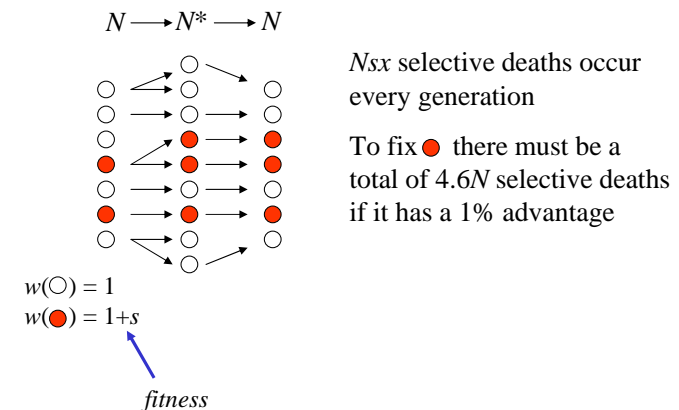
Kimura (1968); King and Jukes (1969)

- Observations
 - Constancy of rate of molecular evolution (the molecular clock)
 - More important regions of proteins evolve at a slower rate than less important domains
 - High levels of protein polymorphism
 - High rates of molecular evolution (about 1.5×10^{-9} changes per amino acid per year)
- Theoretical considerations
 - Haldane's cost of natural selection
 - Segregation load of balanced polymorphisms

Q: Why is there so much variation?

23

Haldane's cost of natural selection

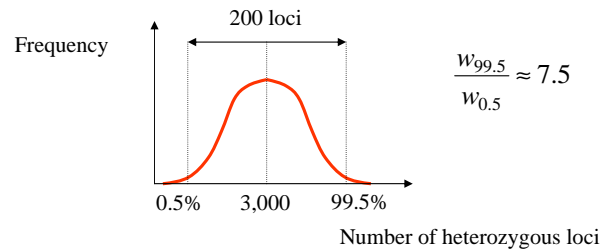


24

The paradox of polymorphism: segregation load

Genotype	AA	Aa	aa
Fitness	$1 - s$	1	$1 - s$
Frequency	x^2	$2x(1-x)$	$(1-x)^2$

15,000 polymorphic loci, maintained by $s = 1\%$



25

Features of the neutral theory

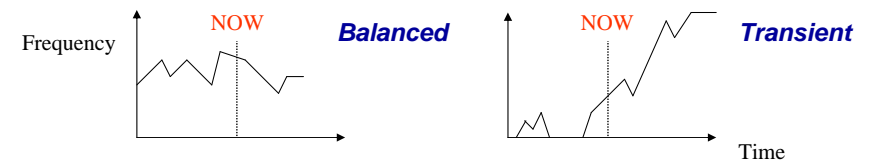
- The majority of changes in proteins and at the level DNA which are fixed between species, or segregate within species, are of no selective importance
- The rate of substitution is equal to the rate of neutral mutation

$$k = f_{neutral} \mu$$

- The level of polymorphism in a population is a function of the effective population size and the neutral mutation rate

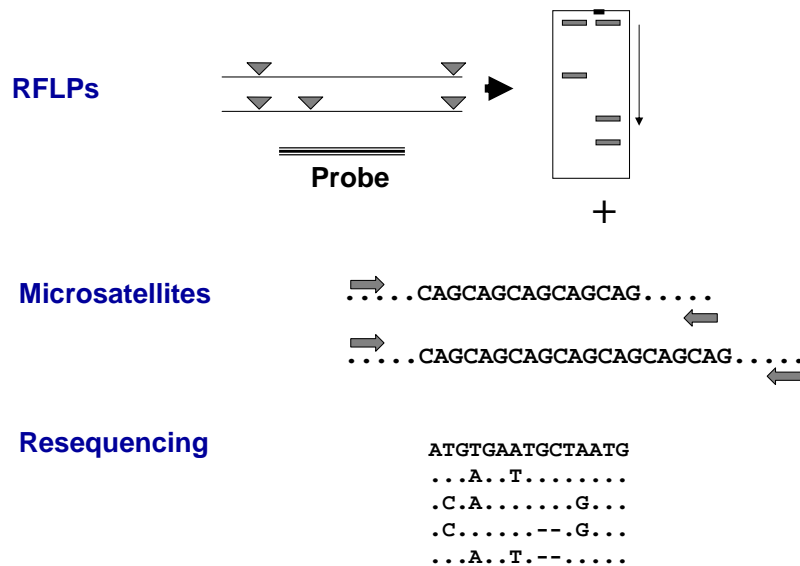
$$\pi = 4N_e \mu$$

- Polymorphisms are transient rather than balanced



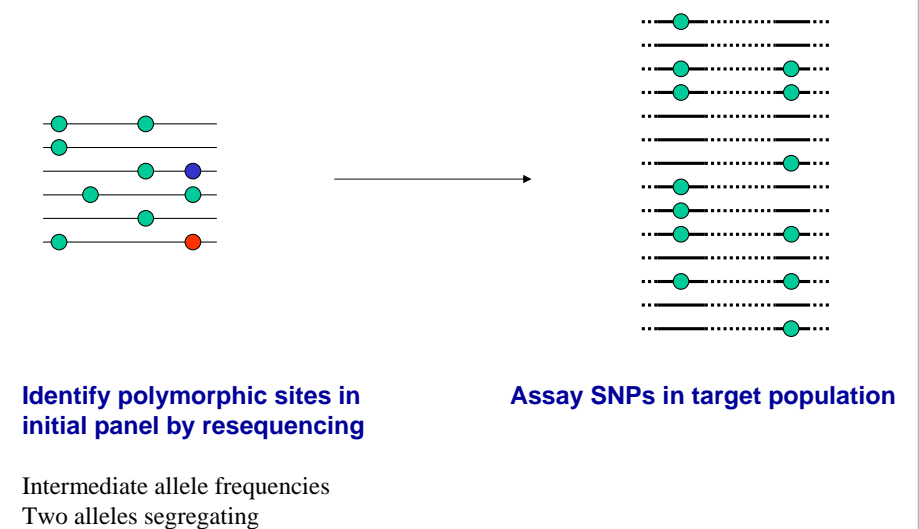
26

Detecting DNA sequence variation



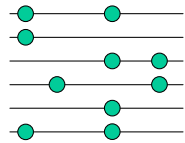
27

SNP analyses (Single Nucleotide Polymorphisms)



28

Statistics of polymorphism



Number of Segregating sites $S = 4$

Average pairwise differences $\pi = 1.9$

Seq	2	3	4	5	6
1	1	2	4	1	0
2		3	3	2	1
3			2	1	2
4				3	4
5					1

Number of distinct haplotypes $H = 5$

29

Patterns of variation at the DNA level

- Synonymous & nonsynonymous mutations

Arg	Gln	Val	Arg	Gln	Val
AGA	CAA	GTA	AGA	CAA	GTA
	↓			↓	
CAG	CGA	GTA	AGA	CAG	GTA
Arg	Arg	Val	Arg	Gln	Val

e.g. *D. simulans*

π_{total}	= 0.010 per site
π_{silent}	= 0.038
$\pi_{\text{noncoding}}$	= 0.023

- Nucleotide variation v. protein variation?

	Humans	<i>D. melanogaster</i>
Allozyme	6%	14%
Nucleotide	0.1%	1%

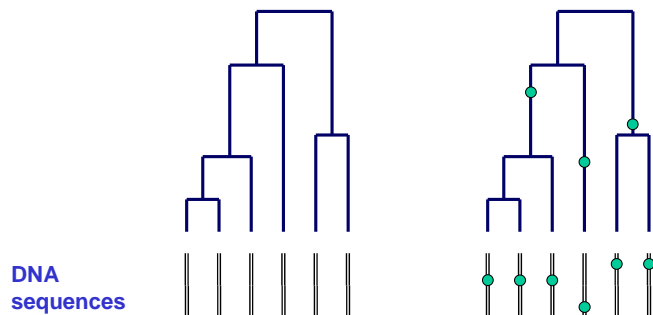
30

Separation of genealogy and mutation

Q: What use are neutral mutations?

Genealogical process

Mutational process



31

Current issues in population genetics

- Medical applications
 - Disease gene identification by association mapping
 - Understanding genetic basis of quantitative variation
- Statistical issues
 - Methods for detecting natural selection
 - Full likelihood methods for estimating evolutionary parameters from sequence data
 - The design of population genetic experiments
- Theoretical and empirical issues
 - The maintenance of quantitative genetic variation
 - Interactions between alleles at selected loci
 - The molecular clock
 - Reproductive isolation and speciation

32