

Classical Selection, Balancing Selection, and Neutral Mutations

Classical Selection Perspective of the Fate of Mutations

- All mutations are **EITHER** beneficial or *deleterious*
 - *Beneficial* mutations are *selected for* and *maintained* in the population
 - **Positive selection**
 - Mutation rapidly increases to a high frequency in the population
 - Generate a new adaptive phenotype
- *Deleterious* mutations are *selected against* and *eliminated* from the population
 - **Negative selection**

Balancing Selection Perspective of the Fate of Mutations

- In general, agrees with Selection Perspective
 - But it was noted that some deleterious variation is maintained in the population
 - How is the deleterious variation maintained?
 - Selection for heterozygotes, one method
 - *Heterozygous* have a *fitness advantage* and undergo *natural selection*
 - Classic example: *human sickle cell anemia*
 - Homozygous normal β -globin allele
 - Proper oxygen carrying capacity, but susceptible to malaria
 - Homozygous mutant β -globin allele
 - Resistant to malaria but die young because of poor oxygen carrying capacity
 - Heterozygous β -globin allele individuals
 - Proper oxygen carrying capacity, and resistant to malaria

Surprise of 1960s and Onward

- *Diversity* in populations ***much greater*** than predicted by either the classical or balancing selection theories
 - Based on protein electrophoresis and eventually sequence data

Neutral Mutations Recognized as a New Class of Mutations

- *Allelic variation neither beneficial or deleterious*
 - *These alleles not provide any fitness difference among individuals in a population*
 - *Fate of an allele in a population is entirely random process*
 - *Allele can be maintained or eliminated*
 - **Controversial concept**

Neutral Theory of Molecular Evolution

Kimura – Nature (1968) 217:624-626

King and Jukes – Science (1969) 164:788-798 (Non-Darwinian Evolution)

Neutral Theory of Molecular Evolution

- Describes the source of variation in natural populations
- The majority of genetic differences between two populations are neutral
 - They have not effect on survival
- Predicts two factors are working
 - Mutation
 - Generates new variation
 - Random genetic drift
 - Fixes variation
 - Stochastic processes that lead to changes in gene frequency
 - Allele can be lost or maintained
- Also called the ***Mutation-Drift Model***

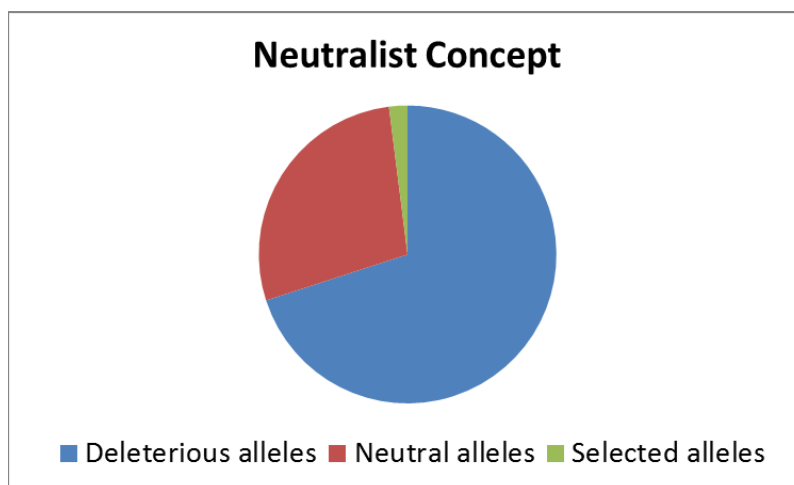
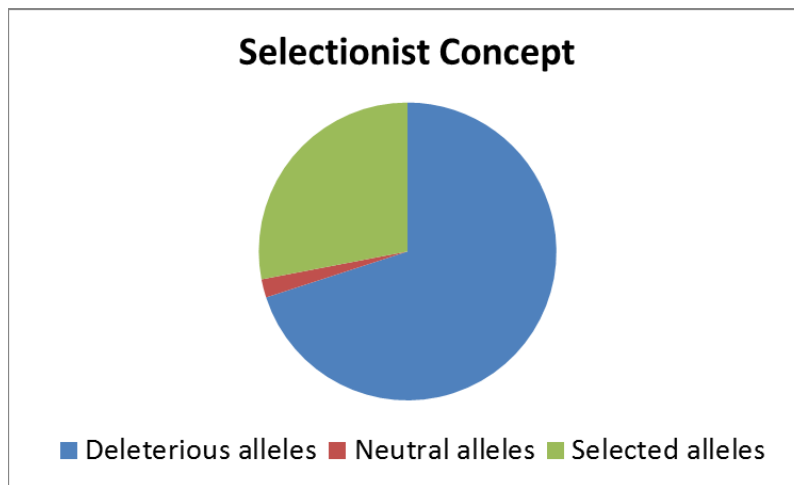
Definition of the Neutral or Mutation-Drift Model

- Population genetic variation results from the appearance of neutral mutations that are fixed (but usually lost) by genetic drift
- Why genetic drift?
 - A process in smaller populations
 - Random process will drive new alleles to fixation quicker

Relationship to Selection Theory

- Most variation is the neutral, therefore *most population differences are not the result of adaptive selection*

Graphical Representation of the Alleles in a Population



The Controversy and Debate

- Was there no role for selection?
 - What is the role (if any) of Darwinian selection for adaptation
 - Although argued, not really a debatable issue
- Some genes do under go selection
- What is the relative distribution of neutral and selected genes

Why is the Neutral Theory so important?

- Ideal for mapping population structure and tracing ancestry
- Provide a null hypothesis for testing for selection

Life Span of an Allele

- Drift will eventually lead to the fixation of one allele in a population
- How long before one allele is fixed (and the other lost)
 - $4N_e$ generations
 - N_e is effective population size
 - The number of individuals, in an idealized ancient population that adheres to the assumptions of Hardy-Weinberg equilibrium, that would evolve to have the same level of diversity as the population that is being observed.
 - Usually less than the census size of population under study

What can affect the life span of an allele?

Balancing Selection

- The maintenance of multiple alleles within a population
- A mechanism
 - Heterozygote advantage
 - Heterozygote has a greater fitness
 - Sickle cell anemia
 - At least two alleles are maintained

Selective Sweep

- A specific gene is the target of selection
- It becomes fixed in the population
- Linked genes become monomorphic (lose variation) during the selection process
 - A direct effect of effect of recombination
 - Hitchhiking
 - Neighboring genes are said to have “hitchhiked”

Background Selection

- Result of eliminating a deleterious allele
 - Entire chromosome carrying deleterious allele is lost
 - Reduces diversity of all genes on that chromosome

Nucleotide Diversity Estimates for a Population of DNA Sequences

Problem 2.4. Principles of Population Genetics; Hartl and Clark; 1997; 3rd Edition
 Gene=*Rh3*; Species=*Drosophila simulans*; gene size=500 nucleotides; sample size=5 lines

Sample	Polymorphic nucleotides															
	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3
f	T	C	T	A	C	C	T	C	C	T	C	G	G	T	T	A
g	T	C	C	T	A	C	C	T	C	C	T	G	G	T	T	T
h	C	T	T	C	C	C	C	T	C	T	T	T	G	C	T	A
i	C	T	T	C	C	C	C	T	T	C	T	G	A	C	T	T
j	C	T	T	C	C	T	C	T	T	T	T	G	G	C	C	A

Pairwise differences

2X3	X	X							X	X					X		X
1X4			X		X	X	X	X				X	X	X		X	
(1X4)+(1X3)				X													

Population Genetic Parameters

S = frequency of polymorphic loci

$$S = \# \text{ polymorphic nucleotides} / \text{total } \# \text{ of nucleotides}$$

Example:

$$S = 16 \text{ polymorphic nucleotides} / 500 \text{ nucleotides}$$

$$= 0.032$$

$\pi = \text{observed average pairwise nucleotide differences in a sample (nucleotide diversity)}$

$$\pi = \frac{\text{sum pairwise differences}}{(\# \text{ pairs})(\text{sequence size})}$$

$$\# \text{ pairs} = \frac{n(n-1)}{2}$$

Example: $n=5$

$$\# \text{ pairs} = \frac{5(5-1)}{2}$$

$$= 10$$

$$\pi = \frac{(6 \times 6) + (9 \times 4) + (1 \times 7)}{(10)(500)}$$

$$= 0.016$$

Variance of π

$$V(\pi) = \frac{b_1 \pi}{k} + b_2 \pi^2 \quad \text{where}$$

$$b_1 = \frac{n+1}{3(n-1)} \quad \text{and}$$

$$b_2 = \frac{2(n^2 + n + 3)}{9n(n-1)}$$

Example: $b_1 = \frac{5+1}{3(5-1)} = 0.5$

$$b_2 = \frac{2(5^2 + 5 + 3)}{(9)(5)(5-1)} = 0.367$$

$$V(\pi) = \frac{(0.5)(0.016)}{500} + (0.367)(0.016^2) = 0.000109$$

Standard Error of Variance of π

$$SE(\pi) = \sqrt{V(\pi)}$$

Example: $V(\pi) = 0.000109$

$$SE(\pi) = \sqrt{0.000109} = 0.0104$$

$\theta = \text{expectation of nucleotide polymorphism}$ in a population when only mutation-drift are occurring (no selection)

$$\theta = \frac{S}{a_1} \quad \text{where}$$

$$a_1 = \sum_{i=1}^{n-1} \frac{1}{i} \quad \text{and}$$

n= sample size

Example: $S = 0.032$

$$a_1 = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = 2.083$$

$$\theta = \frac{0.032}{2.083}$$

$$= 0.015$$

Variance of θ

$$V(\theta) = \frac{\theta}{ka_1} + \frac{a_2(\theta)^2}{a_1^2} \quad \text{where}$$

k = nucleotides in sequence and

$$a_2 = \sum_{i=1}^{n-1} \frac{1}{i^2} \quad \text{and}$$

n = sample size

Example: $k = 500$

$$a_2 = \frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} = 1.42$$

$$V(\theta) = \frac{0.015}{(500)(2.083)} + \frac{(1.42)(0.015)^2}{(2.083)^2}$$
$$= 8.804 \times 10^{-5}$$

Standard Error of Variance of θ

$$SE(\theta) = \sqrt{V(\theta)}$$

Example: $V(\theta) = 8.804 \times 10^{-5}$

$$SE(\theta) = \sqrt{8.804 \times 10^{-5}} = 0.010$$

Tajima's D

- a test for neutrality
- a comparison of π and θ
- under neutrality, the mean of this value is 0
- significant deviation from 0 suggests the gene is undergoing selection

Formula

$$D = \frac{\pi - \theta}{\sqrt{c_1 S + c_2 S(S - \frac{1}{k})}} \quad \text{where}$$

$$c_1 = \frac{b_1}{a_1} - \frac{1}{a_1^2} \quad \text{and}$$

$$c_2 = \left(\frac{1}{a_1^2 + a_2}\right) \left(b_2 - \frac{n+2}{a_1 n} + \frac{a_2}{a_1^2}\right)$$

Example: $c_1 = \frac{0.5}{2.083} - \frac{1}{(2.083)^2}$

$$= 0.0096$$

$$c_2 = \left(\frac{1}{(2.083)^2 + 1.42} \right) \left(0.367 - \frac{5+2}{(2.083)(5)} + \frac{1.42}{(2.083)^2} \right)$$

$$= 0.0038$$

$$D = \frac{0.016 - 0.015}{\sqrt{(0.0096)(0.032) + (0.0038)(0.032)\left(0.032 - \frac{1}{500}\right)}}$$

$$= 0.0567$$

- D not significant different than 0
 - *Physical region or gene evolving via the neutral theory*

What does a significantly **positive** or **negative** Tajima's D value indicate? The following quotes are directly from *A Primer of Population Genetics* (Hartl; 3rd edition):

Significantly positive D value

“The frequencies of polymorphic variants are too nearly equal. This pattern increases the proportion of pairwise differences over its neutral expectation, hence $\pi - S/a_i (= \theta)$ is positive. The finding typically suggests either some type of balancing selection, in which heterozygous genotypes are favored, or some type of diversifying selection, in which genotypes carrying the less common allele are favored.”

- Multiple low frequency alleles are maintained in population
 - Balancing selection at work in the population

Significantly negative D value

“The frequencies of the polymorphic variants are too unequal, with an excess of the most common type and a deficiency of the less common types. This pattern results in a decrease in the proportion of pairwise differences, so $\pi - S/a_i (= \theta)$ is negative. Typical reasons for excessively unequal frequencies are:

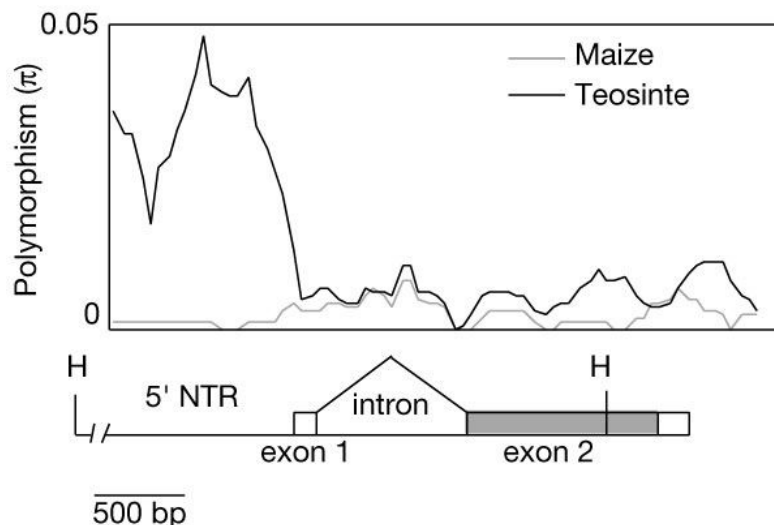
- Selection against genotypes carrying the less frequent alleles
 - Result of a recent bottleneck
 - A few distinct subpopulations appear
 - The bottleneck eliminates less frequent alleles, and insufficient time since the bottleneck to restore the equilibrium between mutation and random drift
 - Selective sweep regions can be observed

The Effects of Domestication on Diversity

Application of population genetics statistics to identify important genetic factors

Maize *tb1* gene studied

- Domestication gene of maize
 - Repressor element expressed higher in maize than teosinte
 - Suppresses branching in maize
- Polymorphism among 13 maize and 9 teosinte compared
 - Coding region
 - Nucleotide diversity (π) was *low* for *both maize and teosinte*
 - 5' non-transcribed region
 - Nucleotide diversity (π) was *much lower* for *maize than teosinte*
 - *Selection acted on the maize 5' non-transcribed region*



Wang et al. (1999) The limits of selection during maize domestication. *Nature* 398:236

Wang et al. (2001) The limits of selection during maize domestication. *Nature* 410:718.

Follow up research on *tb1* 5' region

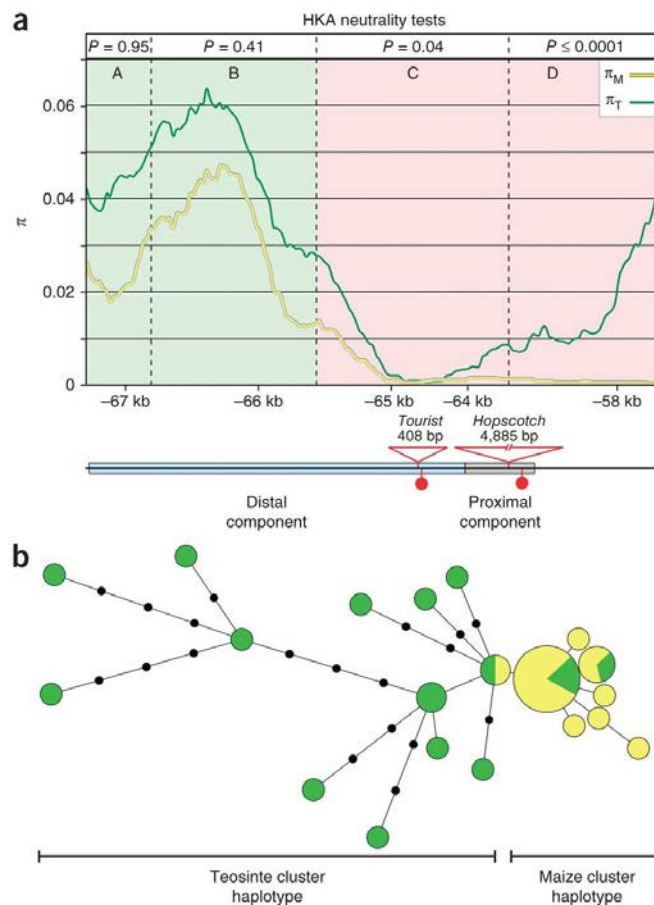
- Low diversity extends to 58.6 kb upstream of gene
- A selection sweep is observed in the upstream region of *tb1* of maize but not teosinte

Locus	Length, bp	Maize			Teosinte		
		n	$\theta \times 10^3$	$\pi \times 10^3$	n	$\theta \times 10^3$	$\pi \times 10^3$
162.9-kb	467	18	10.7	12.3	5	12.1	11.9
93.4-kb	485	14	27.1	20.8	8	38.6	37.5
58.6-kb	520	23	0.5	0.2	—	—	—
45.8-kb	1,003	24	1.1	0.3	9	31.1	32.9
35.6-kb	1,024	24	3.1	1.7	—	—	—
7.1-kb	842	24	6.7	4	8	17.6	12.7
2.5-kb	534	24	3.5	2.8	—	—	—
1.7-kb	935	24	0.6	0.3	8	34.1	34.9
0.4-kb	761	32	3.4	1.4	7	4.6	3.6
5' cDNA	839	32	1.8	1	7	6.8	5.2

Clark et al. (2004) Proceedings of the National Academy of Science 101: 700

Diversity extends to 65 Kb upstream of *tb1*

- Two transposable element are at the border of the low diversity
- *Tourist* element is older
- *Hopscotch* element is new and nearly completely fixed in all maize lines
 - *Is the transposable element the controlling element controlling the increased expression of the *tb1* gene product in maize?*



(a) Nucleotide diversity across the *tb1* upstream control region. Base-pair positions are relative to AGPv2 position 265,745,977 of the maize reference genome sequence. *P* values correspond to HKA neutrality tests for regions A–D, as defined by the dotted lines. Green shading signifies evidence of neutrality, and pink shading signifies regions of non-neutral evolution. Nucleotide diversity (π) for maize (yellow line) and teosinte (green line) were calculated using a 500-bp sliding window with a 25-bp step. The distal and proximal components of the control region with four fixed sequence differences between the most common maize haplotype and teosinte haplotype are shown below. (b) A minimum spanning tree for the control region with 16 diverse maize and 17 diverse teosinte sequences. Size of the circles for each haplotype group (yellow, maize; green, teosinte) is proportional to the number of individuals within that haplotype.

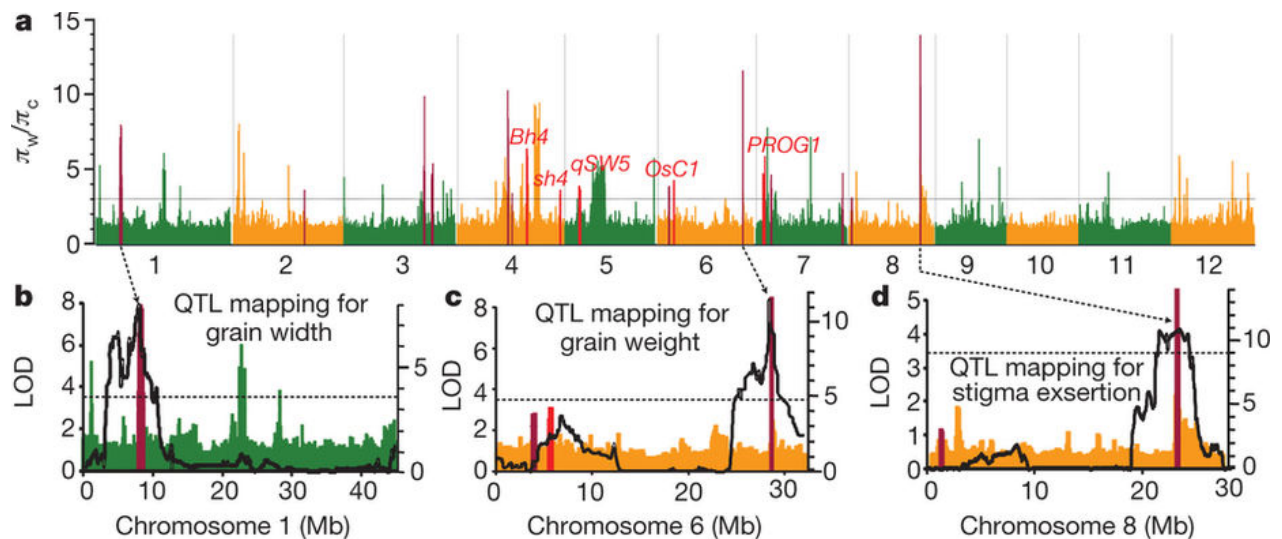
Studer et al. (2011) Nature Genetics 43:1160

Whole Genome Application of Population Genetics Statistics

- The π statistics was estimated across the entire genome of rice
 - Statistic calculated for non-overlapping 100kb regions across all chromosomes
- Diversity (π) between wild (*Oryza rufipogon*) and cultivated (*Oryza sativa*) were compared
 - The following π ratio was used

$$\pi_{\text{wild}}/\pi_{\text{cultivated}} = \pi_{O. \text{ rufipogon}} / \pi_{O. \text{ sativa}}$$

- Regions with high ratios are considered region that under went selection during domestication
 - Why???
 - For domesticated lines, diversity was reduced in region with domestication genes whereas diversity was maintained in wild lines



a, Whole-genome screening of domestication sweeps in the full population of *O. rufipogon* and *O. sativa*. The values of π_w/π_c are plotted against the position on each chromosome. The horizontal dashed line indicates the genome-wide threshold of selection signals ($\pi_w/\pi_c > 3$). **b–d**, A large-scale high-resolution mapping for fifteen domestication-related traits was performed in an *O. rufipogon* \times *O. sativa* population. The domestication sweeps overlapped with characterized domestication-related QTLs are shown in dark red, and the loci with known causal genes are shown in red. Among them, three strong selective sweeps were found to be associated with grain width (**b**), grain weight (**c**) and exerted stigma (**d**), respectively. In **b–d**, the likelihood of odds (LOD) values from the composite interval mapping method are plotted against position on the rice chromosomes. Grey horizontal dashed line indicates the threshold (LOD > 3.5).

How many windows were under selection?

- Cutoff set by permutation test
 - Indica and japonica rice combined
 - 55 windows
 - Indica rice alone
 - 60 windows
 - Japonica rice alone
 - 62 windows

Were domestication genes in the selection windows?

Yes!!!

- *Bh4*: hull color
- *sh4*: seed shattering
- *qSW5*: grain width
- *OsCI*: leaf sheath colour and apiculus colour
- *PROG1*: tiller angle
 - Located in windows with high $\pi_{\text{wild}}/\pi_{\text{cultivated}}$ ratio