The Language of the Universe: Or Using Simple Equations to Teach Powerful Things.
Wilson Memorial Lecture
Research Council on Mathematics Learning
Charlotte, NC March 1, 2019

Dr. Joseph L. Graves Jr.*

Associate Dean for Research & Professor of Biological Sciences

Joint School of Nanosciences & Nanoengineering

North Carolina A&T State University & UNC Greensboro

*Fellow, American Association for the Advancement of Science: Section G: Biological Sciences

2017 BEYA Innovator of the Year

UNC GREENSBORO

Maximizing Access to Research Careers
Undergraduate Student Training in Academic Research

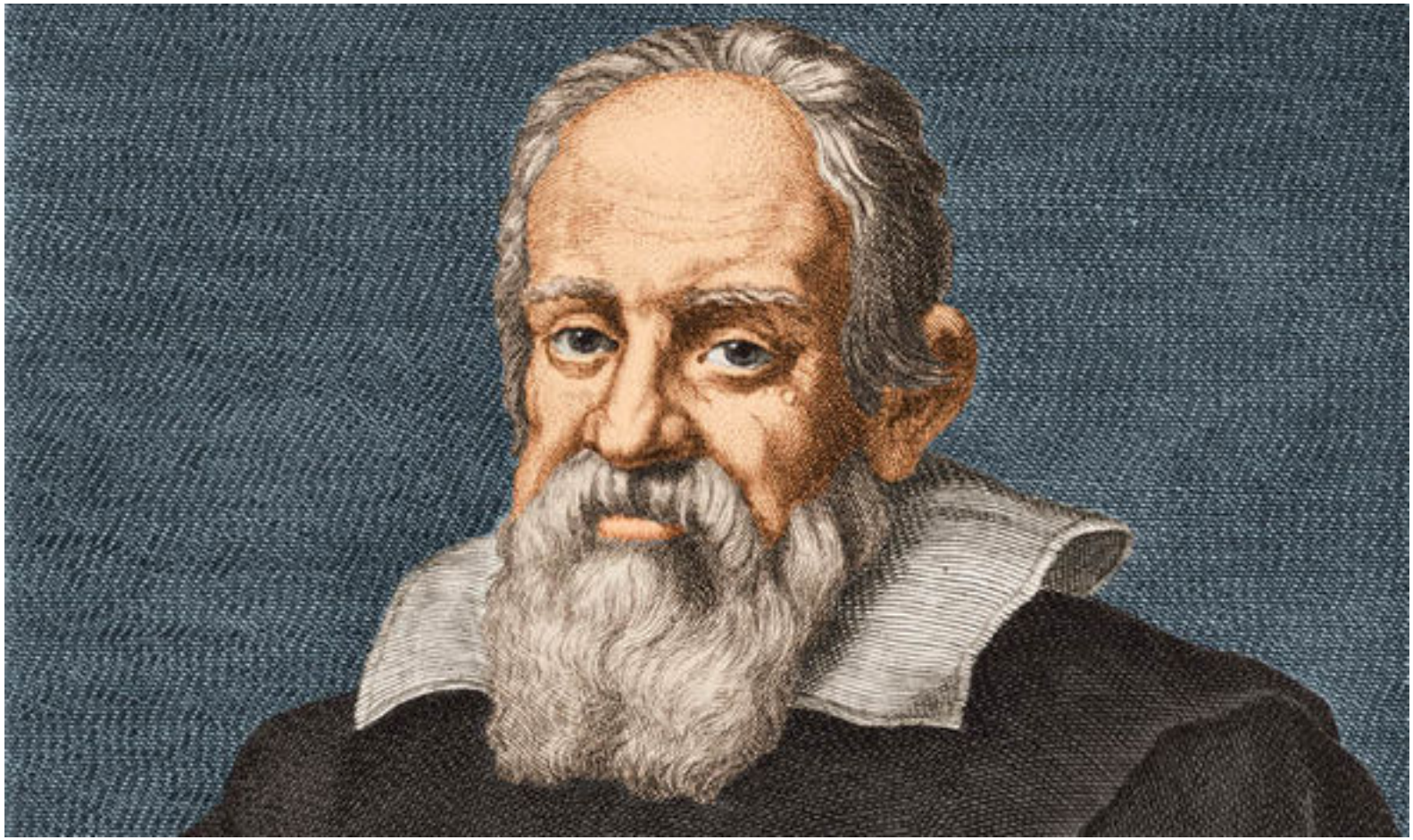
BEACON

SENIC Southeastern Nanotechnology Infrastructure Corridor

IBIEM

Duke GRID

GENOMICS RACE IDENTITY DIFFERENCE



“Mathematics is the language in which God has written the universe”
Galileo Galilei (1547—1642)

“Physics envy is the curse of biology”

- This quote is from Joel Cohen, *Science* vol. 172, May 1971.
- Of course the envy was in part the result of the perceived simplicity of the mathematical formulations of Newtonian physics.
- Yet for modern biology to be birthed it also required a core organizing paradigm with a solid mathematical formulation.
- This begins with the publication of “*On the Origin of Species*” by Charles Darwin in 1859.

The necessity of natural selection

- Three things are requires for natural selection to exist:
- Variation
- Heredity (offspring resemble their parents)
- Struggle for Existence
- The first two chapters concern the reality of variation in biological world (Under Domestication, Under Nature).
- These two chapters also demonstrate the reality of heredity.
- Chapter 3 is the Struggle for Existence.

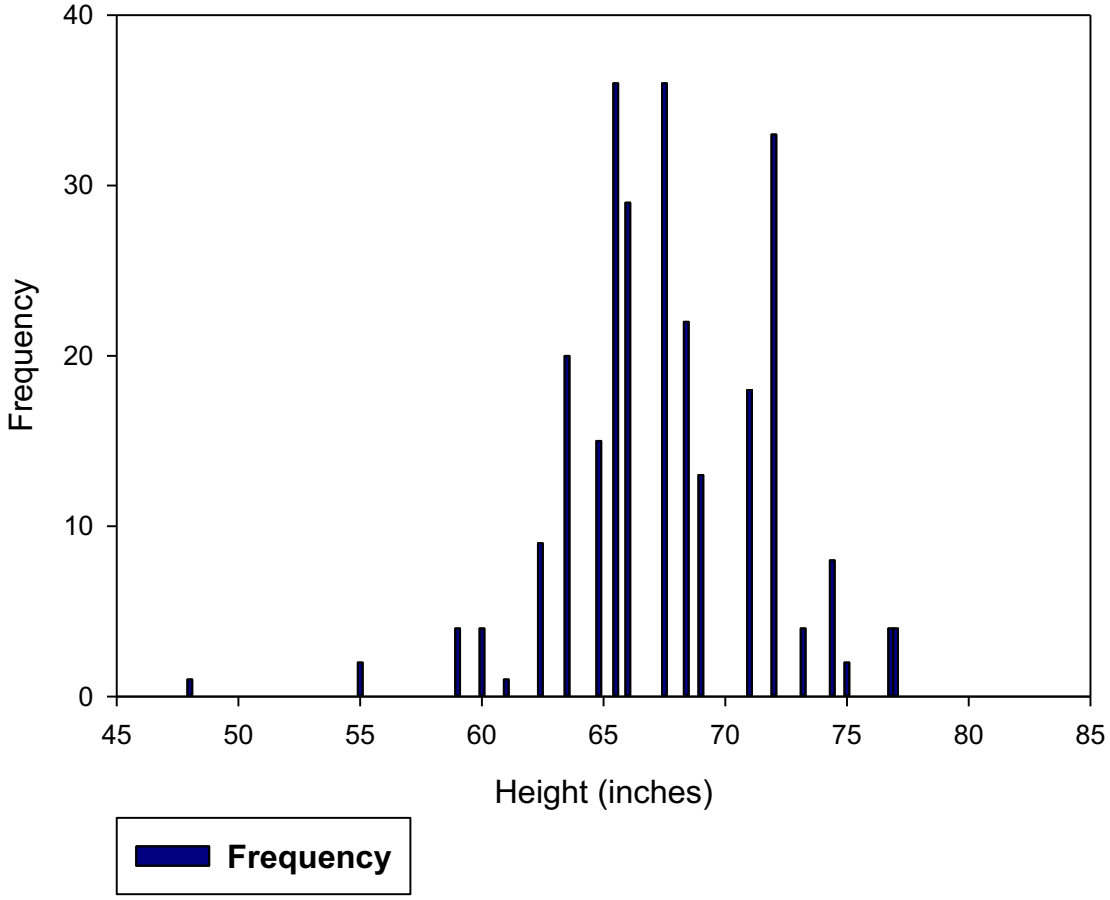
Genetics of complex traits

- Most people are familiar with the genetics of simple or Mendelian traits.
- Despite that, the public labors under many misconceptions concerning even simple genetics.
- Most common is the confusion between traits for which one or a few genes determine much of the outcome versus complex traits (many genes).
- Traits determined by many genes must be measured; hence the term “quantitative genetics.”
- Examples are traits such height, weight, metabolic rate, longevity, or cognitive function.

Mean and variance

- When we measure these traits we usually generate a frequency distribution.
- Often the resulting distributions are normally distributed (Bell Curve).
- These allow us to calculate a mean = $X_b = \Sigma (X_i)/N$; where X_i is the individual observation and N is the sample size. The mean is a measure of central tendency.
- Also we can calculate measures of spread, for example the sample variance = $\Sigma (X_b - X_i)/(N - 1)$

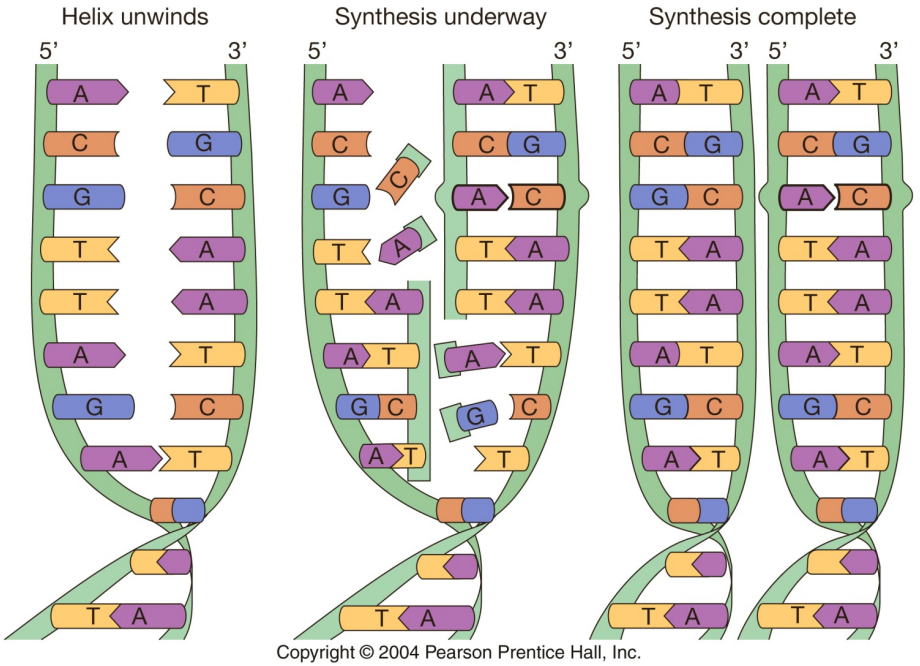
Typical frequency distribution (Quantitative Trait)



Males and females
N = 265.
Mean = 67.3
Variance = 17.3

What about the origin of genetic variation?

- The genetic code of living things on this planet is DNA.
- It is replicated by a series of enzymes called DNA polymerases.
- Each strand of DNA forms a template for the synthesis of the complimentary strand.
- If DNA polymerase inserts the wrong base, it results in a mismatched pair that must be repaired.



Mutations/Genome

Species	**Taxon**	**# of mutations/genome/gen.**	**Genome size**
Escherichia coli	Bacteria*	0.0025	4.6 x 10^6 bp
Sulfolobus acidocaldarius	Archaea*	0.0018	
Neurospora crassa	Fungi*	0.0030	38.6 x 10^6 bp
Saccharomyces cerevisiae	Fungi*	0.0027	12.4 x 10^6 bp
Species	**Taxon**	**# of mutations/genome/gen.**	
Caenorhabditis elegans	Nematode	0.0360	100 x 10^6 bp
Drosophila melanogaster	Insecta	0.1400	122 x 10^6 bp
Mus musculus	Mammal	0.9000	3,400 x 10^6 bp
Homo sapiens	Mammal	1.600	3,300 x 10^6 bp

* single celled organisms that reproduce asexually (via mitosis). Mutation rates seem to be related to the generation time. Prokaryotes, 1 division, fruit fly gamete results from about 25 cell divisions, whereas a human gamete results from about 400 cell divisions.

What factors determine the variance in a complex trait?

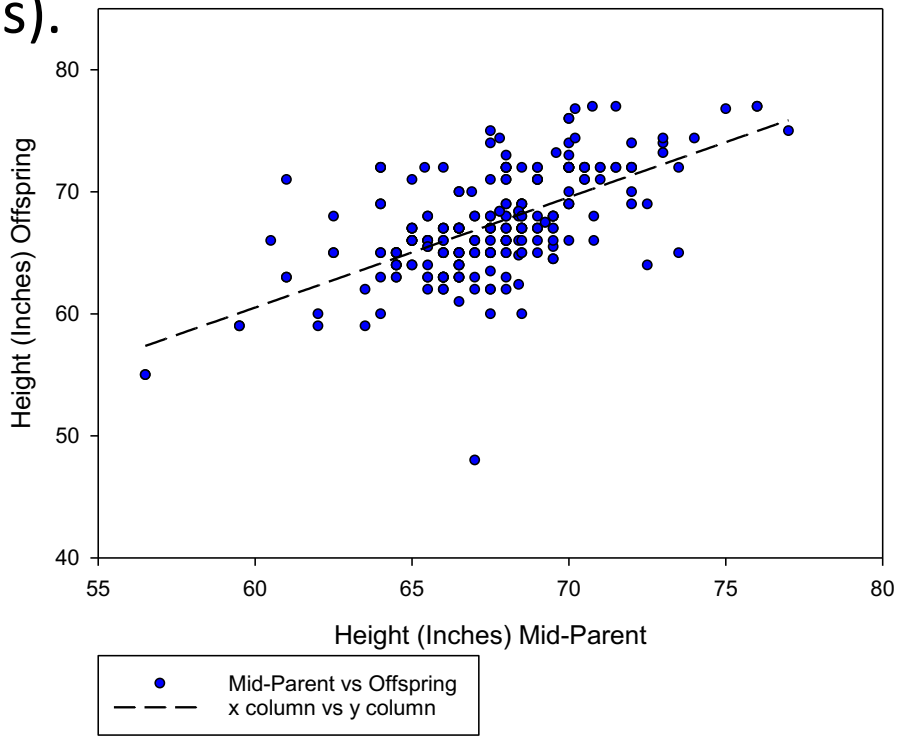
- The variance of any quantitative trait is given by:
- $V_p = V_g + V_e + V_{gxe} + 2Cov\ (G,E) + V_{error}$
- Where V_p is the variance in the trait; V_g variance due to genetic sources; V_e is variance due to environmental sources; V_{gxe} is variance due to gene by environment interaction; Cov (G,E) is the covariance of genes and environment; V_{error} is the error in measuring the trait.

Genetic variance & heritability

- Furthermore $V_g = V_a + V_i + V_d$; where V_a is additive genetic effects; V_i is epistatic effects; and V_d is due to dominance effects.
- However we could also add in epigenetic effects:
- $V_g = V_a + V_i + V_d + V_{ep}$
- Epigenetic are non-nucleotide based changes to the DNA that influence gene expression.
- Heritability (h^2, broad sense) = V_g/V_p
- Heritability (h^2, narrow sense) = V_a/V_p
- From this formula, you should realize that h^2 is limited to go between 0.00—1.00.

Measuring h^2

- Parent – offspring regression
- Concordance between twins (identical and fraternal)
- Full sibling/Half sibling mating designs (cannot be used in humans).



The slope of the regression line is $h^2 = 0.902$.
Well validated studies have Determined h^2 for height in humans at 0.800.

Concordance estimates (selected traits)

Trait	h^2
Blood pressure	0.600
Body mass index	0.500
Verbal aptitude	0.700
Math aptitude	0.300
Spelling aptitude	0.500
General intelligence	0.500—0.800
Longevity	0.100—0.300

Identical (100%)

Fraternal (50%)

Formula $h^2 = (r_i - r_f)/(1 - r_f)$.

Be careful of the smoke and mirrors: You should recognize that estimates of heritability always depend upon the nature of the phenotypic variance; specifically what populations and in what environments where they measured.

Longevity estimates (socially-defined race)

Trait (Longevity)	h2
European Americans (NYC)	0.300
Caribbean Hispanics (NYC)	0.150
African Americans (NYC)	0.100

Lee, Joseph H., Antonia Flaquer, Rosann Costa, Howard Andrews, Peter Cross, Rafael Lantigua, Nicole Schupf, Ming-Xin Tang, and Richard Mayeux. 2004. Genetic influences on life span and survival amongst elderly African Americans, Caribbean Hispanics, and Caucasians. *American Journal of Medical Genetics* 128A: 159–64.

Struggle for existence

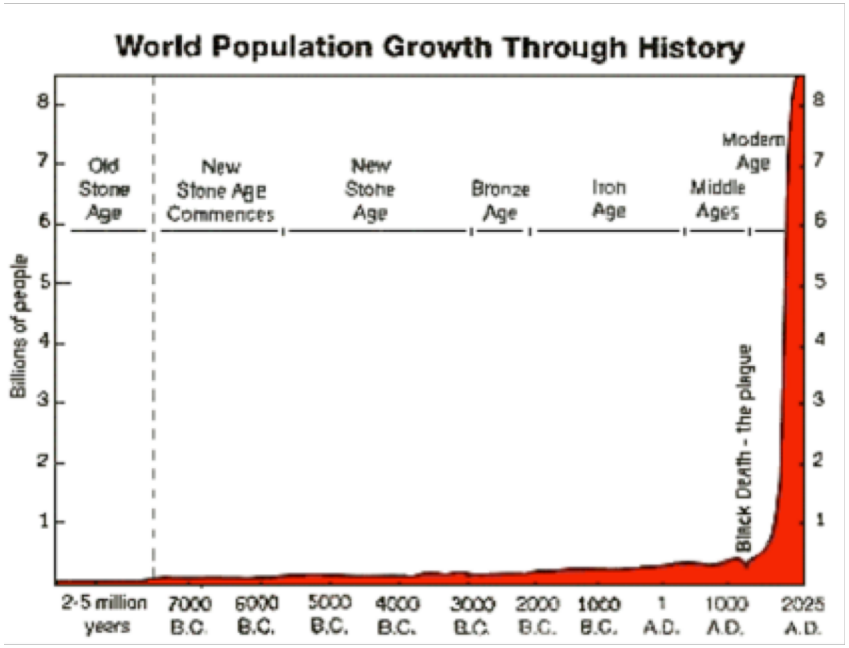
- This follows from the simple fact that living things reproduce.
- Let birth rate = b.
- Let death rate = d.
- We shall define the intrinsic rate of increase as:
- r = b—d
- Now consider that the rate at which a population increases can be calculated as:
- $dN(t)/dt = r * N(t)$
- Where N(t) is the size of the population.

With a little calculus…

- $N_t = N_0 * e^{rT}$
- The NerT equation of exponential growth.

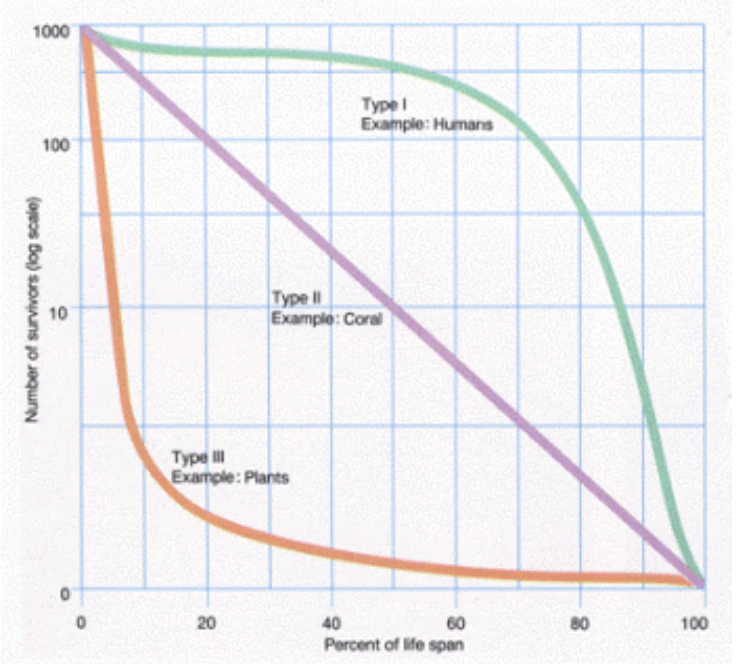
r	T	rT	erT	N_0	N_T
0.02	1	0.02	1.02	10.00	10.20
0.02	2	0.04	1.04	10.20	10.62
0.02	3	0.06	1.06	10.62	11.27
0.02	4	0.08	1.08	11.27	12.21
0.02	5	0.10	1.11	12.21	13.50
0.02	6	0.12	1.13	13.50	15.22
0.02	7	0.14	1.15	15.22	17.51
0.02	8	0.16	1.17	17.51	20.54
0.02	9	0.18	1.20	20.54	24.60
0.02	10	0.20	1.22	24.60	30.04

r	T	rT	erT	N_0	N_T
0.1	18	1.80	6.05	2.67E+08	1.61E+09
0.1	20	2.00	7.39	1.61E+09	1.19E+10
0.1	21	2.10	8.17	1.19E+10	9.74E+10
0.1	22	2.20	9.03	9.74E+10	8.79E+11
0.1	23	2.30	9.97	8.79E+11	8.77E+12
0.1	24	2.40	11.02	8.77E+12	9.67E+13
0.1	25	2.50	12.18	9.67E+13	1.18E+15



Natural selection is not random!
Darwinian or evolutionary fitness

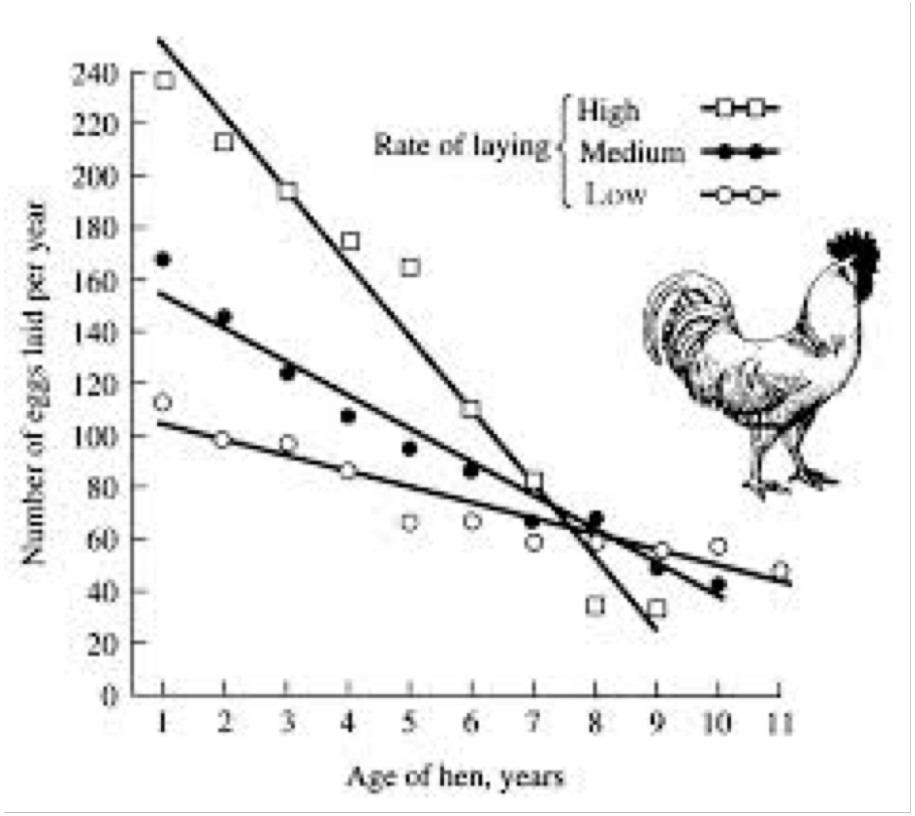
- Do not confuse this concept with the physiological meaning (e.g. physical fitness.)
- This is defined as differential reproductive success.
- This definition has two components: **differential survival** and **differential reproduction**.
- Survival probability is calculated for species with age-structure by determining the number of the original cohort that survives to a given age.
- **This is symbolized as l_x = probability of survival to age x.**
- As a probability its values range from 0.00 to 1.00.



There are 3 general types of survivorship curves.

Differential reproduction

- Even at the same age, individual females will be variable in the number of offspring they produce.
- The **average reproduction (fecundity)** for the female cohort at a given age is symbolized as **m_x.**



fitness = *w* =	**$\Sigma\ l_x * m_x$**	**for ages 0 to X**
l_x =	probability of living to age x	If 1000 flies are born and 100 survive to an age of 45 days, l(45) = 100/1000 = 0.10
m_x =	average reproduction of females at age x.	If 10 female flies aged 45 days produce 20, 30, 15, 10, 4, 0, 21, 25, 3, 30 eggs respectively, m(45) = 148/10 = 14.80
fitness = *w* =	$_S\ l_{45} * m_{45}$	= 0.10 * 14.80 = 1.48

- **Differential reproductive success = fitness = *w***
- **Calculate this from age-specific probability of survival and age-specific reproduction.**

Relative fitness

- Fitness is always determined relative to some other individual or genotype within the population. Consider the following example from *Drosophila pseudoobscura,* in the laboratory, at 21° C:

	l_x Early	l_x Late	m_x Early	m_x Late
PSTRU stock	0.97	0.75	75 eggs	4 eggs
PSTO stock	0.98	0.85	30 eggs	20 eggs

PSTRU stock was produced in laboratory culture with egg collection at 2 weeks.
PSTO stock was produced in laboratory culture with egg collection at 10 weeks.

- Calculating absolute fitness for each stock:
- $W = l_x * m_x$

Absolute Fitness	Early	Late
PSTRU	72.75	3.00
PSTO	29.40	17.00

Now calculate relative fitness for each stock in early and late life.

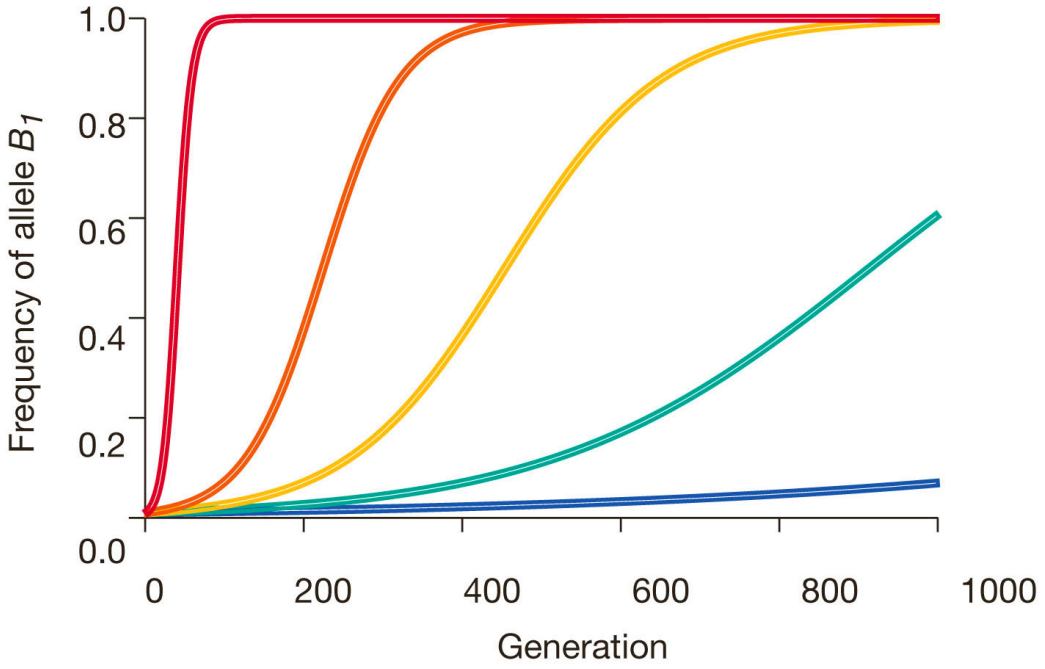
Calculate relative fitness = *w*

Absolute Fitness	Early	Late
PSTRU	72.75	3.00
PSTO	29.40	17.00

- Relative fitness is the comparison of all genotypes to the most fit genotype.
- PSTO_Early = PSTO/PSTRU = 29.40/72.75 = 0.404
- PSTRU_Early = PSTRU/PSTRU = 1.000

Selection coefficient

- We also define **s = the selection coefficient**, such that **$s = 1 - w$.**
- The parameter s is useful if we wish to examine the change in allele frequency due to natural selection.
- For example we can derive the following equation for selection against a recessive gene:
- **Selection against recessive: $\Delta q = spq^2/(1 - sq^2)$**



Selection scheme

	Percent surviving		
	B_1B_1	B_1B_2	B_2B_2
Strong			
	100	90.0	80.0
	100	98.0	96.0
	100	99.0	98.0
	100	99.5	99.0
	100	99.8	99.6
Weak			

- The results above show equation 2 in action.
- In this case you are looking at ΔP (remember as q goes down, p must go up!)
- This is because 1.00 = p + q or p = 1 – q.

Simulation of selection equation

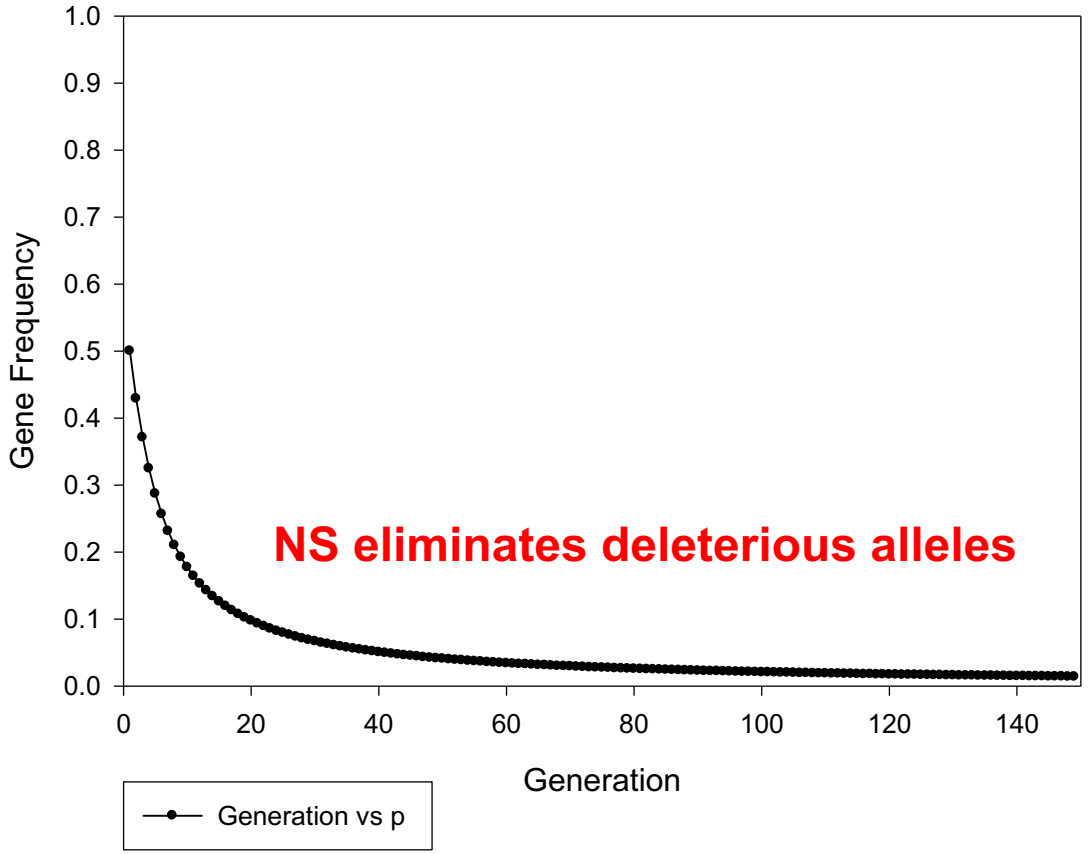
s	q	p	q2	(1-sq2)	spq2	dq	q'	p'
0.900	0.990	0.010	0.980	0.118	0.009	0.075	0.915	0.085
0.900	0.915	0.085	0.838	0.246	0.064	0.260	0.656	0.344
0.900	0.656	0.344	0.430	0.613	0.133	0.217	0.438	0.562
0.900	0.438	0.562	0.192	0.827	0.097	0.117	0.321	0.679
0.900	0.321	0.679	0.103	0.907	0.063	0.069	0.252	0.748
0.900	0.252	0.748	0.063	0.943	0.043	0.045	0.206	0.794
0.900	0.206	0.794	0.043	0.962	0.030	0.032	0.175	0.825

- You can simulate the selection equation in Excel using an iterative calculation (over and over again.)
- This is an example of strong selection against the recessive allele.

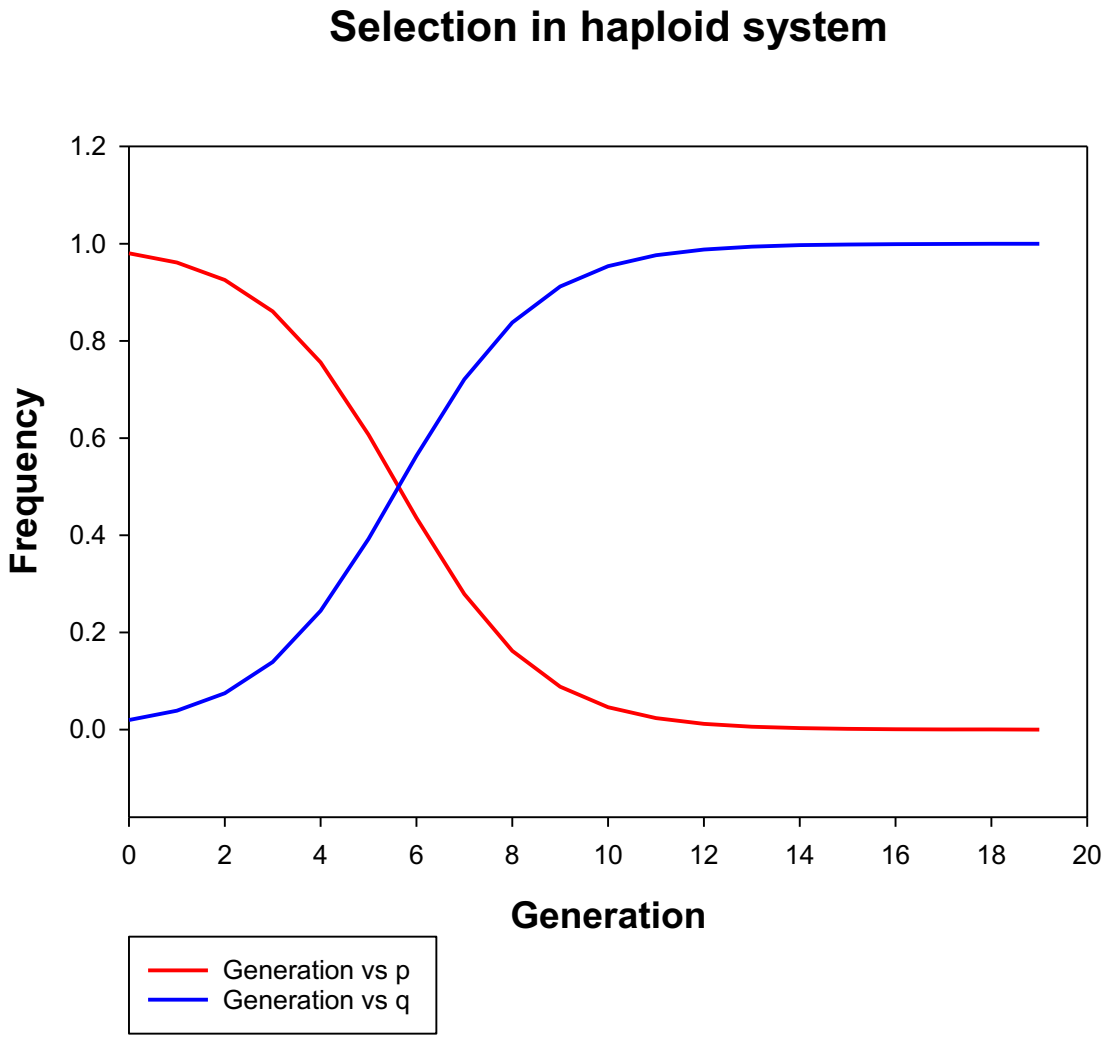
Selection script

- Simpler than using Excel you can write a selection program using any scripting language, such as Python, Matlab, or R.
- First you will need a text editor, for Mac I recommend Text Wrangler (https://www.barebones.com/products/textwrangler/); however as this product is being phased out, probably better is BBEdit: (https://www.barebones.com/products/bbedit/).
- For IBM systems I recommend Notepad++ (https://notepad-plus-plus.org/).
- A useful support text to help you learn the basics is: Haddock and Dunn, Practical Computing for Biologists, Sinauer, 2011).

Fates of an Allele Under Simple Negative Selection



- **Simulation of equation 2, with p = 0.5, and s = 0.90**
- **This simulation shows that the recessive allele essentially goes to extinction.**



	A_1	A_2
Frequency	p	q
Rel. Fitness	w	1.00

- Average fitness; $W_{avg} = pw + q$
- $p_{t+1} = pw/W_{avg}$
- Iterating across generations gives the curve above.

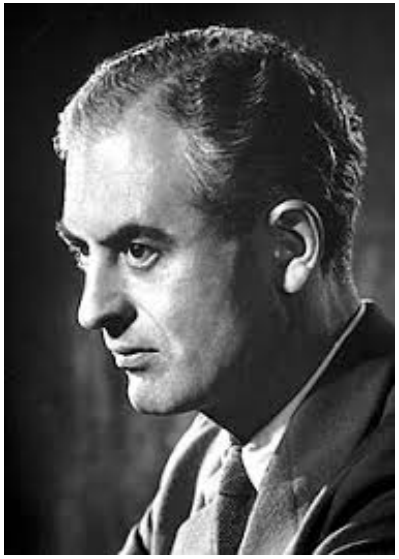
General model of population genetics (script)

- Using the general model of population genetics model, write a script that allows you to calculate allele frequency changes in a population.
- You should be able to input initial allele frequencies, fitness values, and immigration and emmigration of alleles into the population.
- I recommend using Python, but any other language can be used.

Step 1	Determine Genotype frequencies at Birth
Step 2	**Calculate selection by differential survival and reproduction (also immigration/emigration.)**
Step 3	Arrive at new *Genotype Frequency among adults*.

Medawar's Theory

'The force of natural selection weakens with increasing age – even in a theoretically immortal population, provided that it is exposed to real hazards of mortality. If a genetical disaster (...) happens late enough in individual life, its consequence may be completely unimportant.'



$$v(x) = \int_x^\infty e^{-rx} l(t) m(t) d\,t e^{rx} / l(x)$$

Sir R.A. Fisher's equation defining the reproductive output of an individual alive at age (X), in terms of the proportion of the growth of the population as a whole. *Genetical Theory of Natural Selection*, 1930.

Peter Medawar **b. 1915– d. 1987**

Evolution, 38(5), 1984, pp. 1004–1010

LABORATORY EVOLUTION OF POSTPONED SENESCENCE IN *DROSOPHILA MELANOGASTER*

MICHAEL R. ROSE
Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

Received May 24, 1983. Revised December 31, 1983

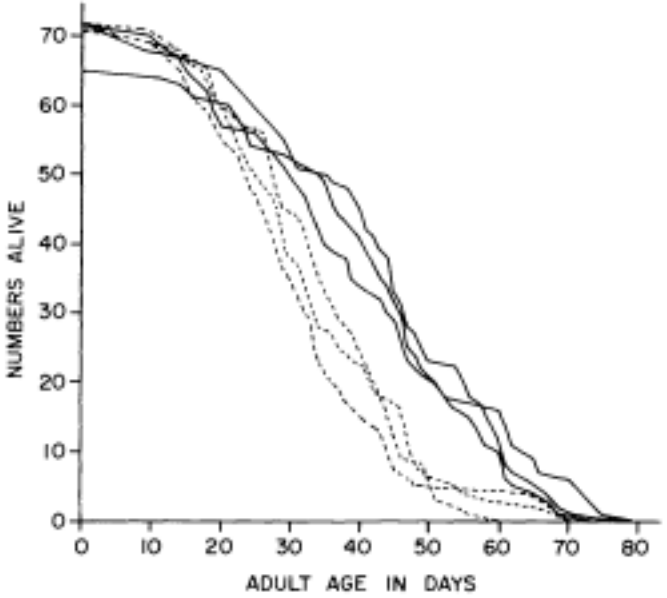


FIG. 1. Surviving numbers of females from the start of the adult life-history assay period. B population samples are shown as dashed lines, O samples as solid lines.

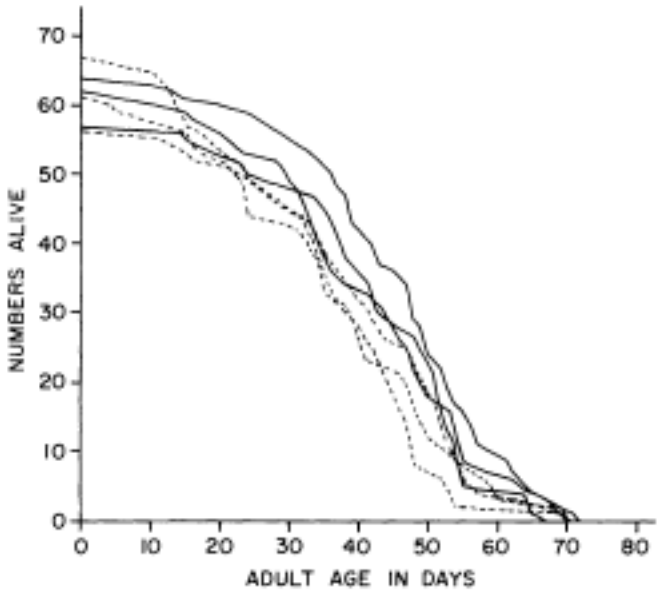
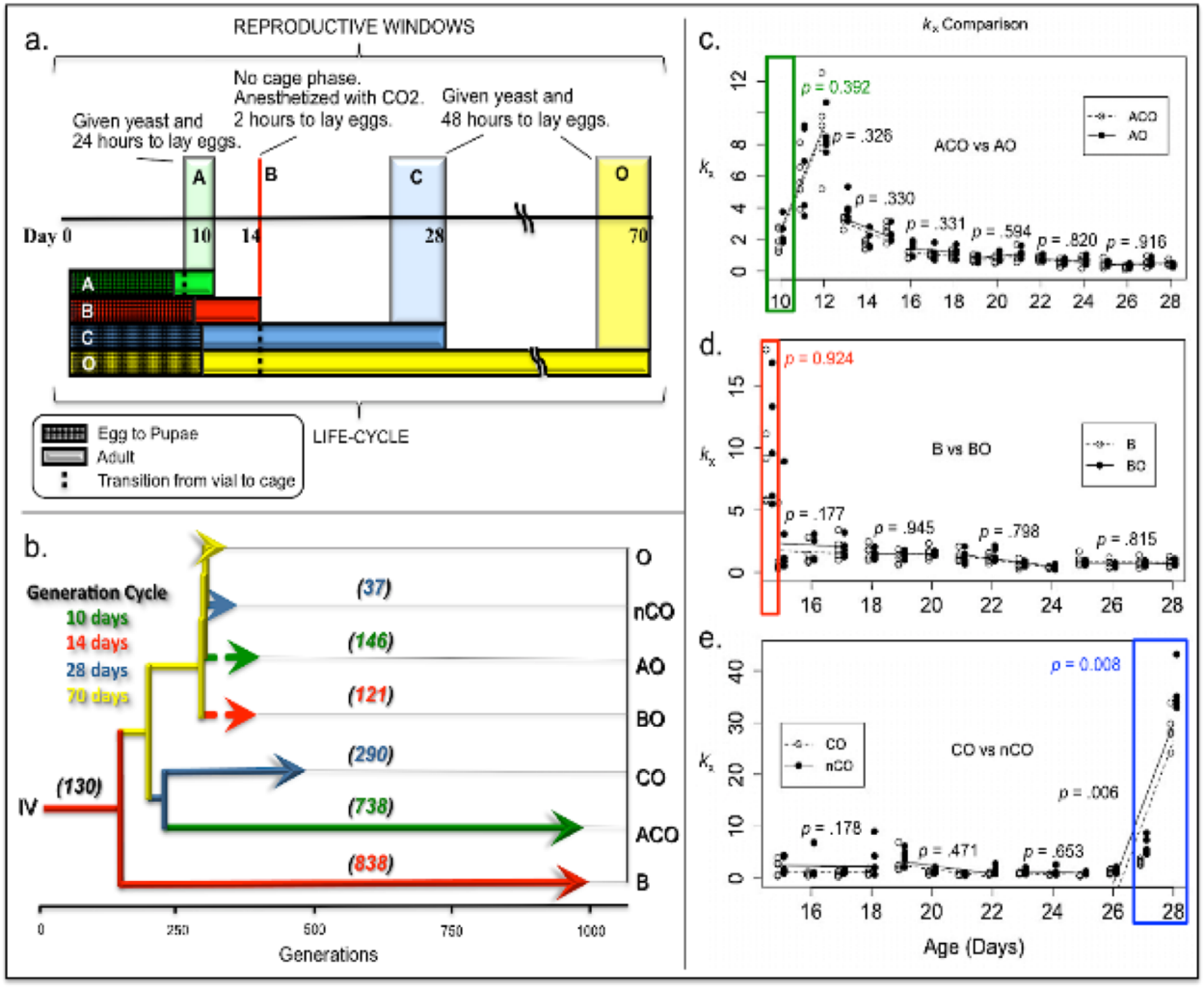


FIG. 2. Surviving number of males from the start of the adult life-history assay period. B population samples are shown as dashed lines, O samples as solid lines.

Joseph L. Graves Jr: How Repeatable is Evolution Genome-Wide?

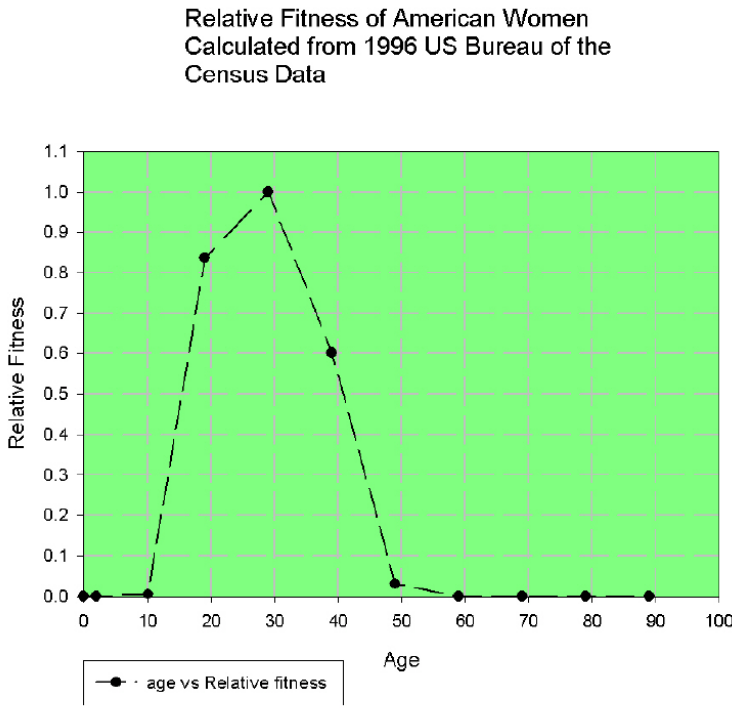


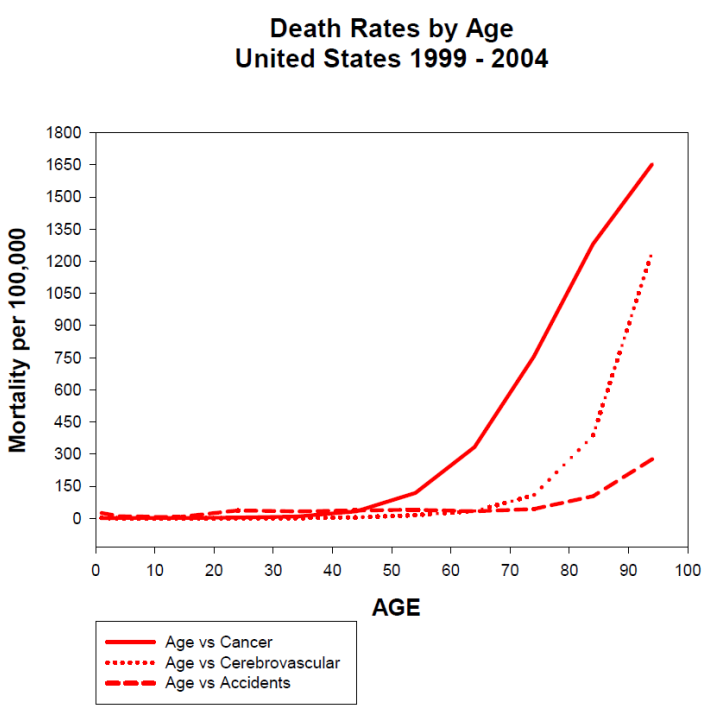
Graves, J.L., Hertweck, K.L., Phillips, M.A.,, Han, M.V., … and Rose, M.R. 2017. Deep Genomics of Convergent Experimental Evolution in *Drosophila*, *Molecular Biology and Evolution*, *doi: 10.1093/molbev/msw282.* First published online: January 12, 2017.

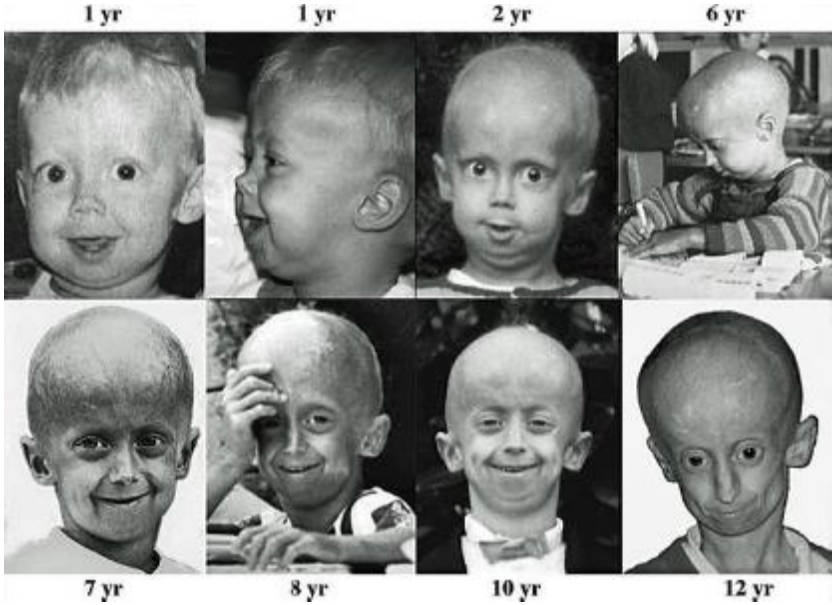
Population genetic mechanisms allowing the evolution of aging.

Mechanism	Early	Late	Senescence
Mutation/selection balance	-	-/+/0	No
Mutation accumulation	0	-	Yes
Antagonistic pleiotropy	+	-	Yes

Examples of diseases associated with mutation accumulation: Alzheimer's Disease; Antagonistic pleiotropy: cancer.







Progeria

Genotype	Fitness (*w*)
AA, Aa	(1 - s)
aa	1.0

The only way A can exist is via mutation. That rate is μ.
Total loss rate of A allele = p * s. At the equilibrium frequency; the rate of gain of the A gene = rate of loss of the A gene: [μ * (1 - p)] = (p * s); Solving gives us this equation: p' = μ/(s + μ).
Given that the selection coefficient >> greater than the mutation rate, we can estimate the last equation as: **p' = μ/s** This is the origin of the term: **mutation/selection balance.**

Study	Frequency	Country
Epstein et al. (1966)	[1 - 22.1]/1,000,000	World
Fraccaro et al. (1966)	[2.2 - 10.8]/1,000,000	Sardinia
Goto et al. (1985)	[2.5 - 3.3]/1,000,000	Japan

Making Sense: Hypothetico-deductive method

- **Nothing in Biology Makes Sense Except in the Light of Evolution,** Theodosius Dobzhansky (1973). *The American Biology Teacher, 35*(3), 125-129.

- Over the course of the early 20th century the Neo-Darwinian synthesis (Mendelian genetics/Natural selection) produced an extensive mathematical theory of selection and genetic drift.
- In the late 20th century the mathematics of the neutral theory of molecular evolution was added to this arsenal.
- These tools allow us to formulate powerful and testable hypotheses in evolutionary biology.
- This in turn has allowed us to unify all of the biological sciences under one enduring paradigm.