05/1/22

BI 3433 - EVOLUTION

Lecture 01

Variation - A Central Theme in Biology

"Nothing in Biology makes sense except in the light of evolution" - Dobzhansky

Every biologist works with variation in one form or another.

Questions asked about it -

1) What is that nature of variation and what is its doing?
2) How is it being generated?
   Its not just genes anymore - epigenetics & environment
3) How is it being transmitted?
   Traditionally studied in genetics
4) How does it change?
   This is classically the domain of evolutionary biology
   1. Within a group (Population & Quantitative Geneticists)
      a) Processes - various interactions (ecology)
      b) Patterns - what happens due to those interactions
         (selection, drift, mutation, migration) + transmission
         affects the rate of production of variation & patterns they cause
   2. Leading to formation of new groups?
      a) Processes : speciation - there was an assumption that processes in 4.1.a cut over a long time to result in speciation - but there might be other, different processes.
      b) Pattern : phylogeny
5) Practical applications - breeding, antibiotic resistance etc

This is a conceptual roadmap of all of biology.

What is Evolution?

Transformation of species through time, including both changes with species and origin of new species.

Two digitally different aims -
- To explain and understand the history of life on earth
- It includes studying the processes acting today - predicting what effects they will have

Three ideas of species diversity
- Creationism
- Transformism (Lamarck)
- Darwinism : descent with modification.

What is a species?
Contested definition. We create arbitrary boundaries to categorise living beings.

* Morphological species concept
Defined by phenotypic characteristics. It applies to both sexual and asexual species.

* Biological species concept
- Species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups (Mays 1940)
- Doesn't apply to fossils or asexual organisms

* Phylogenetic species concept
- Species is an irreducible group whose members are descended from a common ancestor and who all possess of a combination of certain traits.
- Applies to asexual species. but it can be difficult to determine the degree of difference required for separate species which leads to a 'lot' of species.

The concept of 'change' (phenotypic trait, reproductive isolation, genetic similarity) is different across three definitions. So, the kind of evidence needed will also be somewhat different

For Darwin, species were defined by morphological traits and he explained the origin of different traits through natural selection.
This explanation won't work well enough if you change the definition of species.

Changes occur in traits of animals - peppered moth in England, different breeds of dogs etc.
If we consider evolution as 'change' and species as 'sufficiently different groups' then this amount of difference strongly suggests that evolution happens.
But if we consider BSC definition, then its not enough. Can variations within species be sufficiently large to produce reproductive isolation?

1/2/22

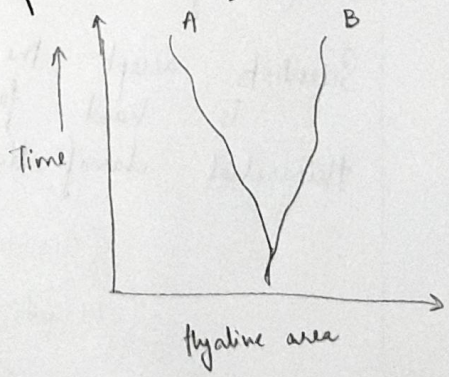
Lecture 2

Industrial melanism and dog breeds are examples of significant change but they can't be called different species. So is this evidence enough?

Rig. species - a continuum of related species around a geographic divide where adjacent species can mate with each other except for the two species at the end

Eg: Salamanders in California.
This challenges BSC, but not Darwin's idea.

Fossil evidence for species transformation
Analysis of diatom cell wall (hyaline cell wall) shows that as we go back in time, the overlap of the trait increases



④ → Can new species be created from old ones?

Yes. Artificially, Primula kewensis from 2 types of primrose by chemically treating the zygote to double the number of chromosomes.

Naturally occuring - wheat, groleysis sp.

So these processes over a long time can give rise to new species & explain diversity.

Does this mean all the existing species came to be this way? Not necessarily. *

So we go to common ancestry.

Analogy - structural similarity explainable by common function
Eg: streamlined body of whale & shark

Darwinian homology / Unity of Type - structural similarity not explainable by functional similarity
Eg: Forelimbs of vertebrates, Genetic code

Explanations for (near) universal genetic code —

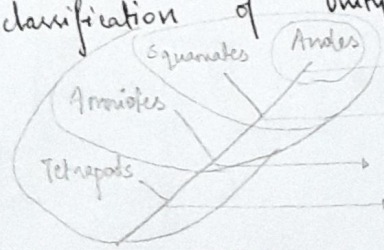
1. Chemical constraints (now found)
   * Some mutations can make UAG t-RNA bind to amino acid instead of stopping translation
   * No steric constraints either.

2. Convergent evolution
   • Owing to robustness of mistranslation, the most optimal code has been selected. Slightly controversial, but most believe (through simulations) that this is not the most optimal

3. Common ancestry
   Frozen accident seems to be the simplest solution (although retaining the possibility of some interaction with others)

Scientists accept the most parsimonious solution, but this is hard for others to accept.

Heirarchal classification of "Unity of type" is possible



Getting a pattern is not remarkable by itself. But if we take consider any set of such traits, we get the same (or similar) pattern again and again.

Penny et al – 11 species, 5 proteins, we get some ~15 trees out of 34 million possible ones. Getting this by chance is vanishingly small. So the simplest solution is common ancestry.

Fossil records also correspond with the phylogenetic tree.

Other evidences : vestigial organs, pseudogenes and primitive forms during development

But all of this requires you to accept parsimony (as a good way of accepting answers) to accept common ancestry as explanation.

Existence of transitional forms

Homology of bones of jaw in reptiles, mammals and animals in between. Reptiles have 4 bones, mammals have one bone in jaw & 3 others in the inner ear. The fossils of intermediate animals show transition states

To summarize –

1. Evidence for origin of new species is direct and uncontroversial
2. Each strand of evidence for common ancestry, by itself, has to appeal to parsimony.
3. Taken together, they make a very strong case for believing in common ancestry.
4. The last statement is the reason for which the debate lingers.

Lecture 3

2/2/22

Modelling Evolution – Population genetics

Starting at the lowest level

Evolution is a change in frequency of alleles in the gene pool over time.

Transgenerational model of microevolution make assumption about the mechanism of –

1) producing gametes
2) uniting gametes
3) developing phenotypes

1) Producing gametes – we need to specify the genetic architecture –
   - number of loci and their genomic positions
   - no. of alleles per locus
   - rate of mutation
   - mode and rules of inheritence of genetic elements (here is where MES and EES shools differ).

2) Uniting gametes (for sexual reproduction) – weneed Population Structure –
   - system of mating (random vs non-random)
   - size of population
   - presence, amount and pattern of genetic exchange (migration etc.)
   - age structure of individuals in population : it affects the effective size of population

3) Developing phenotypes – more imp for quantitative genetics
   It can range from no GxE interaction to different types of GxE interactions – epistasis.

   phenotype = Gene x Environment

Hardy-Weinberg Model

Assumptions:

- One autosomal locus
- Two alleles
- No mutation
- Mendel's first law
- Random mating
- Infinite population
- One isolated population
- No age structure
- All genotypes have identical phenotypes wrt. their ability for replicating DNA.

This is a null case. Individuals are essentially gene pairs. Transmission of genes is abstracted into sampling from a bagful of gametes.

One locus . Two Allele Scenario

A_1 , A_2 alleles are present in frequencies p , q

Frequency of genotype – A_1A_1 - P A_1A_2 - Q A_2A_2 - R

Then. $P + Q + R = 1$

$p + q = 1$

$p = P + \frac{Q}{2}$ $q = R + \frac{Q}{2}$

How to solve problems – ~21 mins

Deriving Hardy-Weinberg Law

Freq of A_1 in Parents — p

↓ Normal segregation, Equal fertility, No mutation, No migration

Freq of A_1 in gametes — p

↓ Large population, Equal fertilizing capacity of all gametes

* Important assumptions

Freq of A_1 in gametes forming zygotes — p

↓ Random mating, Equal freq in males & females

Genotypic freq among zygotes — p^2 $2pq$ q^2

↓ Equal viability

Genotypic freq among the progeny — p^2 $2pq$ q^2

↓ Equal viability

Freq of A_1 in progeny

$p = p^2 + \frac{1}{2} \cdot 2pq$

$= p(p+q)$, $(p+q) \to 1$

$= p$

Statement

" In a large mating population with no selection, mutation or migration, the gene frequencies and genotypic frequencies are constant from generation to generation, and there's a simple relation between gene frequencies and genotypic frequencies "

(8) H-W law is a null law. It played an important role in clarifying the nature of inheritance and existence of variation.

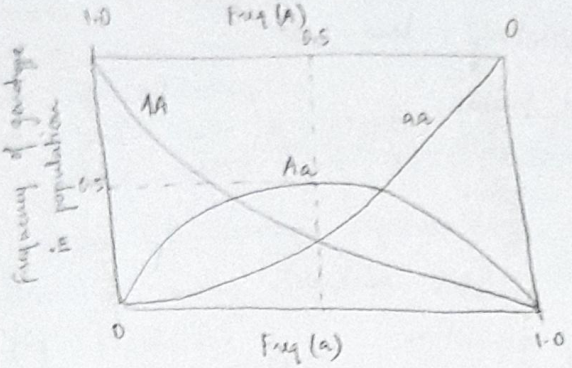
Darwin's theory of blending inheritance would end up in no variation at all. Directed mutagenesis is not an explanation either.

So how is variation maintained in the face of selection?

This null law showed that if there's no selection, variation is maintained in the population, and not lost.

Population genetics is built on deviations from H-W law.

We can also arrive at H-W law by writing down mating frequencies - Weinberg's paper (table)



max (Freq(Aa)) = 0.5

If freq(a) is low, it'll mainly occur as heterozygotes

Points about H-W law -

1. Relationship b/w gene and genotype freq is reached in a single generation.
2. F2 genotype freq (P', Q', R') only depend on p, q and not P, Q, R (genotypic freq) of F1
3. max (freq (Aa)) ≤ 0.5 & this occurs when p = q = 0.5
4. When freq (A) is low, the rare allele occurs as heterozygote and dominant allele occurs as homozygote
5. Its in neutrally stable equilibrium - if you perturb, that will become new equilibrium.

8/2/22

Lecture 4

Extension to 1-locus multi-allelic case : $(p+q+\dots n)^2$

Extension to X-linked inheritance

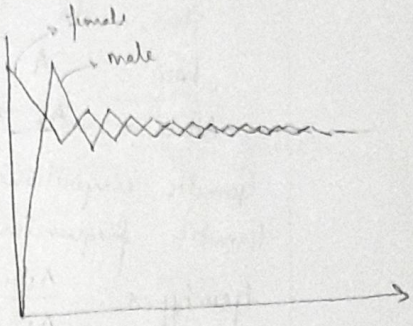
Females ↓ / Males →	X_A	X_a	Y
	p_m	$1-p_m$	
X_A p_f	X_AX_A $p_m \cdot p_f$	$X_A X_a$ $(1-p_m)\, p_f$	X_AY p_f
X_a $(1-p_f)$	X_AX_a $(1-p_f)p_m$	X_aX_a $(1-p_m)(1-p_f)$	X_aY $(1-p_f)$

Alleles : A, a

Males in next generation only get the allele from their mother. So, $p_m' = p_f$

$$p_f' = X_AX_A + \frac{1}{2}(X_AX_a + X_aX_A)$$

$$= \frac{1}{2}(p_m + p_f)$$



$$p_m' - p_f' = \frac{1}{2}(p_m - p_f) \qquad -\frac{1}{2}(p_m - p_f)$$

The freq among males & females oscillates i.e. it's more in one sex every other generation.

Even if frequencies are different to begin with, they'll converge soon to a common frequency.

Overall freq of A (p') in next generation will be a weighted average of A across the sexes –

$$p' = \frac{1}{3}p_m' + \frac{2}{3}p_f' = \frac{1}{3}\left(p_f' + 2\cdot\frac{1}{2}(p_f' + p_m)\right) = \frac{1}{3}(2p_f + p_m)$$

⟹ Deviations from HW equilibrium

1. Small population sizes leading to drift
2. Mutation
3. Migration
4. Selection
5. Non-random mating
   - Assortative mating
   - Dis-assortative mating
   - Inbreeding
   - Population sub-structure.

How to detect deviations from HWE?

Chi-squared test

Say : $N_{AA} = 20$ $N_{Aa} = 10$ $N_{aa} = 10$ degrees of freedom = 1

$P = 0.5$ $Q = 0.25$ $R = 0.25$

$p = 0.625$ $q = 0.375$

∵ if we know p, everything else is fixed.

Compute the test statistic : $\chi^2 = \sum \frac{(O-E)^2}{E}$

O : Observed
E : Expected freq

	Observed	Expected	$\frac{(O-E)^2}{E}$
AA	20	15.6	1.2
Aa	10	18.75	4.08
aa	10	5.6	3.4
		Σ	= 8.71

Compare the test statistic against the table value for the appropriate df and significance level.

If computed χ^2 value is greater than table value, then reject the null hypothesis.

Two-locus case

Here, it's easier to follow the gametes than the alleles

Loci : A , B

Allele : A_1, A_2 ; B_1, B_2

Gametic compositions : A_1B_1 A_1B_2 A_2B_1 A_2B_1

Gametic frequencies : x_{11} x_{12} x_{21} x_{22}

$\sum_{ij} x_{ij} = 1$

Genotypes : $\frac{A_1B_1}{A_1B_1}$ $\frac{A_1B_1}{A_1B_2}$ $\frac{A_1B_1}{A_2B_1}$ $\frac{A_1B_1}{A_2B_2}$ $\frac{A_1B_2}{A_1B_2}$ $\frac{A_1B_2}{A_2B_1}$ $\frac{A_1B_2}{A_2B_2}$

$\frac{A_2B_1}{A_2B_1}$ $\frac{A_2B_1}{A_2B_2}$ $\frac{A_2B_2}{A_2B_2}$

$\Rightarrow$ x_{11}^2 $2x_{11}x_{12}$ $2x_{11}x_{21}$ $2x_{11}x_{22}$ x_{12}^2 $2x_{12}x_{21}$ $2x_{12}x_{22}$

x_{21}^2 $2x_{21}x_{22}$ x_{22}^2

Frequency in the next generation : $\frac{A_1B_1}{A_2B_2}$ — gives A_1B_1 when there's no recombination

$$x_{11}' = x_{11}^2 + \frac{1}{2}2x_{11}x_{12} + \frac{1}{2}\cdot 2\cdot x_{11}x_{21} + \frac{1}{2}\cdot 2x_{11}x_{22}(1-r) + \frac{1}{2}\cdot 2x_{12}x_{21}\cdot r$$

$$x_{11}' = x_{11}(x_{11} + x_{12} + x_{21}) + x_{11}x_{22} - r\cdot x_{11}x_{22} + r\cdot x_{12}x_{21}$$

$x_{11} + x_{12} + x_{21} + x_{22} = 1$

$$= x_{11}(1) - r(x_{11}x_{22} - x_{12}x_{21})$$

$$x_{11}' = x_{11} - rD \quad \text{where} \quad D = x_{11}x_{22} - x_{12}x_{21}$$

Gametic disequilibrium measures the degree to which
→ the frequencies of different gametic types differ from what is predicted based on allele freq alone

If $D = 0$, then $x_{11}' = x_{11}$ but usually $D \neq 0$, then its called Gametic / linkage disequilibrium

Frequency in next gen:

$$x_{11}' = x_{11} - rD \qquad x_{12}' = x_{12} + rD$$
$$x_{22}' = x_{22} - rD \qquad x_{21}' = x_{21} + rD$$

★ Frequency of $A_1 \to P_1$ Here, $P_1 + P_2 = 1$

$B_1 \to q_1$ $q_1 + q_2 = 1$

$$P_1 = x_{11} + x_{12} \qquad q_1 = x_{11} + x_{21}$$
$$P_2 = x_{21} + x_{22} \qquad q_2 = x_{12} + x_{22}$$

$$P_1' = x_{11}' + x_{12}' = x_{11} - rD + x_{12} + rD$$
$$= x_{11} + x_{12}$$
$$\therefore P_1' = P_1$$: Allele frequency is not changing!

Genotypic / gametic frequencies maybe changing, yet there's no change at the allele level.

★ Disequilibrium in next gen

$$D' = x_{11}' x_{22}' - x_{12}' x_{21}'$$
$$= (x_{11} - rD)(x_{22} - rD) - [(x_{12} + rD)(x_{21} + rD)]$$
$$= x_{11}x_{22} - rD(x_{11} + x_{22}) + r^2D^2 - [x_{12}x_{21} + rD(x_{12} + x_{21}) + r^2D^2]$$
$$= x_{11}x_{22} - x_{12}x_{21} - rD(1) = D - rD$$

$$D' = D(1 - r)$$

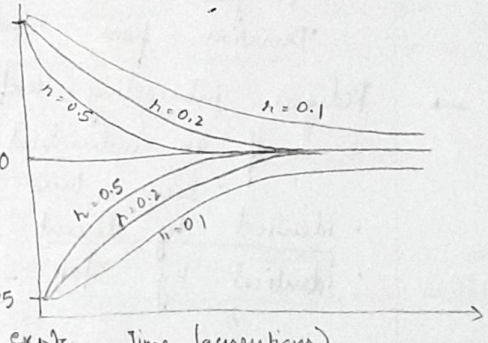
Max value of $r = 0.5$

$$D_t = D_0(1 - r)^t$$

Gametic disequilibrium decays at a steady rate determined by r.

⇒ Even loci on different chromosomes do not go into equilibrium in a single generation — it takes time.

Imp to keep it mind while doing popgen



Lec 5 - 9/2/22

Two locus extension

Gametic equilibrium : two locus gamete frequencies are the products of the constituent single locus allele frequencies i.e. knowing one locus in a gamete doesn't alter the probabilities of alleles in second locus ie. they are independent

Gametic disequilibrium : departure from equilibrium.

Its given by.

$$D = x_{11} - p_1 q_1$$

p_1 : freq (A_1)

q_1 : freq (B_1)

$$p_1 q_1 = (x_{11} + x_{12})(x_{11} + x_{21}) = x_{11}^2 + x_{11}x_{21} + x_{11}x_{12} + x_{12}x_{21}$$

$$= x_{11}(x_{11} + x_{12} + x_{21}) + x_{12}x_{21}$$

$$= x_{11}(1 - x_{22}) + x_{12}x_{21} = x_{11} - x_{11}x_{22} + x_{12}x_{21}$$

$$p_1 q_1 = x_{11} - (x_{11}x_{22} - x_{12}x_{21})$$

$$\therefore p_1 q_1 = x_{11} - D$$

Similarly. $D = x_{22} - p_2 q_2$ $\quad D = x_{12} + p_1 q_2$ $\quad D = x_{21} + p_2 q_1$

Single value of D can be used to compute disequilibrium in any combination.

Recombination and linkage disequilibrium are sufficient conditions for evolution in a multilocus system.

15/2/22

Lecture 6

Systems of mating

We relax the assumption of random mating.

Inbreeding : colloquially. mating between relatives. We'll look at two things -

Pedigree inbreeding coefficient

Deviation from random mating expectations.

→ Pedigree inbreeding coefficient

If an individual has 2 identical alleles at a given locus. then it can be due to :

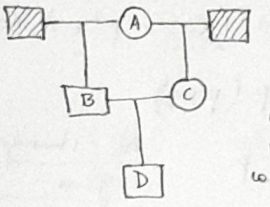
- identical by descent - both are copies of same ancestral allele
- identical by state - different origins

If we go far back enough, everyone is related. So, we choose an arbitrary reference population, for which its assumed that everyone is unrelated, so alleles are identical by state.

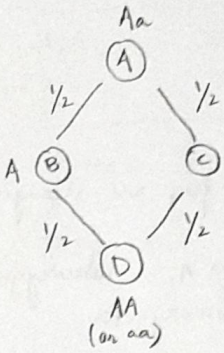
Pedigree inbreeding coefficient (F) is defined as probability that a focal individual in a pedigree is homozygous due to identity by descent at randomly chosen locus.

$0 < F < 1$ — if $F > 0$, organism is inbred

★ Example :- Half-siblings mating



Simplify by excluding inds who don't contribute to IBD



Fathers in F_1 have alleles unrelated to the mother, so they don't contribute to IBD

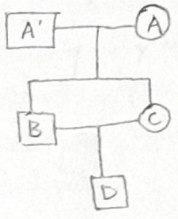
This calculation works if F_1 mother is homozygote also, as long as they not identical by descent.

$$P(D = AA) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

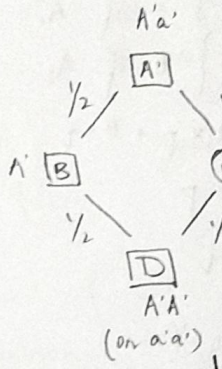
$$P(D = aa \text{ or } AA) = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}$$

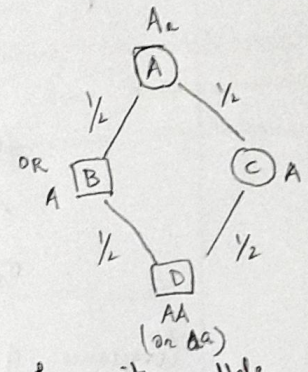
$\Rightarrow$ For ind D, $F = 1/8$

★ Full siblings mating



Simplify by splitting into two mutually exclusive loops that can contribute to IBD





An individual can be homozygous by descent for either allele from only one of common maternal/paternal ancestors. So, total prob. of ~~IBD~~ IBD, $F = 1/8 + 1/8 = 1/4$ for D

Used to be worked out by Catholic church to give dispensations for incestuous marriages, way before Mendel!

F is applied to specific individuals, not population, though there can be a distribution of F's in the population. It depends critically on the state of the founding population.

Inbreeding as deviation from HWE

Here, we mean that prob. of fertilisation of A_1 with A_2 is not the same as that of A_1. So we expect different gametic frequencies in the population

λ : deviation parameter from simple product of gamete frequencies

homozygotes come together more frequently than heterozygotes. If we don't use $+\lambda$ & $-\lambda$, then we go into the realm of selection.

		A_1	A_2
		P_1	P_2
A_1	P	A_1A_1 $p^2+\lambda$	A_1A_2 $pq-\lambda$
A_2	q	A_1A_2 $pq-\lambda$	A_2A_2 $q^2+\lambda$

$$\begin{bmatrix} P = p^2+\lambda \\ Q = 2pq-2\lambda \\ R = q^2+\lambda \end{bmatrix}$$

$$p' = P + \tfrac{1}{2}Q = p^2+\lambda+pq-\lambda$$

$$p' = p(p+q)$$

Only genotypic freq. are changing. $\therefore p' = p$: No change in gene frequencies

If $\lambda < 0$, $A_1 \leftrightarrow A_2$ heterozygotes come together more often than homozygotes.

We need to find variance and covariance of ? ~32-33 mins

We define random variables X and Y of for males and females such that,

$$X = \begin{cases} 1 & \text{if } A_1 \\ 0 & \text{if } A_2 \end{cases} \qquad Y = \begin{cases} 1 & \text{if } A_1 \\ 0 & \text{if } A_2 \end{cases}$$

$\sigma^2 = \sum (x_i - \bar{x})^2$ f

$$\mu_x = 1\times p + 0\times q$$

$$\mu_x = p$$

$$\sigma_x^2 = pq$$

$$\mu_Y = p$$

$$\sigma_Y^2 = (1-p)^2 \cdot p + (0-p)^2 q$$

$$= q^2p + p^2q = pq(p+q)$$

$$\sigma_Y^2 = pq$$

Covariance of X & Y

$(cov(x,y) = \Sigma (x_i - \bar{x})^a (y_i - \bar{y})^b$ f

$$cov(x,y) = (1-p)(1-p)(p^2+\lambda) + (1-p)(0-p)\cdot 2(pq-\lambda) + (0-p)(0-p)(q^2+\lambda)$$

$$cov(x,y) = q^2(p^2+\lambda) - 2\cdot pq(pq-\lambda) + p^2(q^2+\lambda)$$

$$= p^2q^2 + q^2\lambda - 2p^2q^2 + 2pq\lambda + p^2q^2 + p^2\lambda = \lambda(p^2+2pq+q^2)$$

$$\therefore \boxed{cov(x,y) = \lambda}$$

λ can vary from $-\infty$ to $+\infty$, to constrain, if we divide by $\sigma_x \times \sigma_y$ $\quad f \in [-1, 1]$

Pearson's correlation coefficient –

$$f = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\lambda}{\sqrt{pq}\sqrt{pq}}$$

$$\therefore \boxed{f = \frac{\lambda}{pq}} \Rightarrow \lambda = pq \cdot f$$

Consider $\quad R = 2pq - 2\lambda = 2(pq - pq \cdot f)$

$$R = 2pq(1-f)$$

Exp (hetero) $= 2pq$ $\qquad$ Obs (hetero) $= 2pq(1-f)$

$$\frac{\text{Obs}}{\text{Exp}} = 1 - f \Rightarrow \boxed{f = 1 - \frac{\text{Obs}}{\text{Exp}}}$$

$f > 0 \Rightarrow$ Obs < Exp : there's loss of heterozygosity

F and f are not similar, there are some differences –

Important differences

F	f
• Pedigree data for specific individuals are required	• Genotype freq. data for specific locus in population
• Probability : $0 \leq F \leq 1$	• Correlation coefficient : $-1 \leq f \leq 1$
• For individual	• For population
• Expected probability of identity by descent at randomly chosen autosomal locus for certain ind. caused by biological relatedness of individual's parents	• System of mating measured as deviations from random-mating genotype frequency expectations.

For same population, F and f can lead to different interpretations. See Templeton book. (ch 3).

22/2/2022

Lecture

Assortative mating

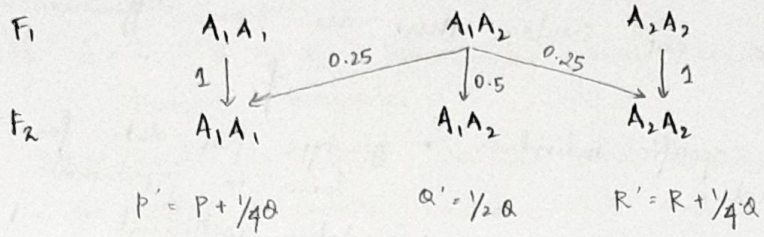
Individuals with similar phenotypes are mating with each other at a greater probability than expected.

This leads to positive correlation between trait values for mating pairs in population. Its not always due to mating preferance.

Eg: Treehoppers – moths that develop on different host plants
hosts determine the rate of development and females have a narrow receptive window, so they tend to mate with males from same plant.
But when other males are available, preferance disappears.

Complete assortative mating (extreme form)

1-locus 2-allele scenarios



$P' = P + \frac{1}{4}Q$ $\quad Q' = \frac{1}{2}Q$ $\quad R' = R + \frac{1}{4}Q$

Eventually, heterozygotes will disappear. But allele frequency doesn't change –

$$p' = P' + \frac{1}{2}Q' = P + \frac{1}{4}Q + \frac{1}{2}\cdot\frac{1}{2}Q = P + \frac{1}{2}Q = p$$

When homozygosity increases, it also leads to expression of diseases that are caused by rare, recessive alleles

Eg: Profound ~~Early Onset Disease~~ Deafness (EOD) in humans

Caused by genetic conditions (68% of cases). There are 115 genes & mtDNA variants, all are rare except for one –
35delG mutation at GJB2 locus: freq is 0.01 in US & Europe

Expected phenotypic freq: 1 in 10,000
Observed: 3-5 in 10,000

This is because there is assortative mating among deaf people – >80% in US and >90% in Europe.

But there are complications

Impact of assortative mating in yeilding an 'f' is proportional to —
a) phenotypic correlation b/w mates
b) correlation b/w genotype & phenotype — not very high for deafness coz it could be caused by accident / disease

¼ of people with EOD are homozygous for recessive allele of GJB2 locus. (Only the children of these couples) ×

Only category of deaf couples to have deaf children are marriages b/w two recessive homozygous individuals

Expected : $\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$ ⇒ 1 in 16 couples should produce deaf children

But. the actual figure is 1 in 6. Why?

→ 100% assortative mating for phenotype determined by two-loci

Alleles at 2 loci — A, a and B, b such that capital letters contribute +1 & small letters 0. Phenotypic values would be like —

AA BB - 4
Aa Bb - 2
aa BB - 2
AA Bb - 3
aa bb - 0
Aa bb - 1

Say we have 100% assortative mating where matings with same phenotipic values are allowed

So, only | 18 | 55 | matings are allowed

AA BB & aa bb can only mate with individuals of same offspring genotype

When other mating pairs produces either of these phenotypes genotypes. then the offspring can't mate with anyone else

So these two states act as absorbing states

~~AA~~ B. ① AA bb × AA bb
② AA bb × aa BB
③ aa BB × aa BB

} There 3 mating pairs produce only one type of offspring.

1, 3 - same as parents

2 - genotype different from parents can never become absorbing state.

(18)

(1) or (3) can becomes absorbing state if only one of them is present as for phenotype '2'. (?) - Confirm

The 4 possible equilibrium states we get are —

AA BB aa bb AA bb aa BB

Among these, AAbb and aaBB cannot coexist, only one of them will persist at equilibrium, depending on their initial freq.

Let allele frequencies be P_A, P_a, P_B and P_b (What is const? $P_A + P_a = ?$)

Genotypes	$P_A = P_B$	$P_A < P_B$	$P_A > P_B$
AA BB	P_A	P_A	P_B
AA bb	0	0	$P_A - P_B$
aa BB	0	$P_B - P_A$	0
aa bb	P_b	P_b	P_a

No change in allele frequency, but phenotypes completely disappear
⇒ Gamete types are also disappearing ⇒ clear linkage disequilibrium
Assortative mating (AM) leads to linkage disequilibrium
Also, there's clear increase in homozygosity at expense of heterozygosity

General insights from a simple case

1. AM increases homozygotes
2. Multiple equilibria are possible and final outcome is decided by intial frequency. (Another factor contributing to historicity of evolution)
3. AM can create and maintain linkage disequilibrium
4. AM causes gametes to bear alleles at different loci that cause similar phenotypic effect (brings together like alleles and separates unlike ones)
5. AM can split a population into genetically differentiated and isolated subsets. Its a potent force for microevolution and speciation

(?) in 45 mins

AM and EDD (Vona et al 2014)

They found that avg. no. of deleterious variants in 80 genes -

Deaf : 3.7 Control : 1.4

Deaf individuals have 2.6 more mutations, recessivity at any loci out of several could be causing deafness.

AM causes linkage disequilibrium and increases homozygosity which together explains why autosomal recessive deafness occurs far more frequently in humans than expected under random mating.

23/2/22

Lecture

Assortative mating vs Inbreeding

In 1-locus scenario, neither can lead to allelic frequencies

In multi-locus, 100% AM / inbreeding, we'll see how they differ.

Recal in 2-locus AM case, homozygotes (4 types) produce offspring that are identical to their parents, so they act as absorbing states

Important differences

	100% Selfing	AM
Absorbing state	4	2 or 3
All gametes present at equilibrium	Yes (depends on itial condtn)	No
Can this create LD on its own?	No	Yes
Can this maintain LD?	Under some special circumstance of 100% selfing	Yes
Which loci get affected?	All (true for inbreeding in general)	Loci related to phenotype that affects mating and other loci at LD with it

Disassortative mating

preferential mating of individuals with dissimilar phenotypes

Eg: Mice show disassortative mating for MHC (since it causes odour variation). Human females are known to prefer dissimilar MHCs from males (sweat expt) Wedekind et al 1995

Say 1-locus-2-allele system with 100% disassortative mating.

A - p a - q

AA - P Aa - Q aa - R

	AA	Aa	aa	Mating freq
AA × aa		1		PR
AA × Aa	½	½		PQ
Aa × aa		½	½	QR

$$P' = \frac{1}{2}\frac{PQ}{SUM} \qquad Q' = PR + \frac{1}{2}(PQ + QR)/SUM \qquad R' = \frac{1}{2}\cdot\frac{QR}{SUM}$$

These frequencies won't sum to 1. So we scale them by

$$SUM = PQ + RP + QR$$

Consider, $p = 0.25 \qquad q = 0.75$

$P = 0.0625 \qquad Q = 0.375 \qquad R = 0.5625 \qquad p = 0.25$

$P' = 0.04 \qquad Q' = 0.565 \qquad R' = 0.39 \qquad p' = 0.326$

1. ⇒ Heterozygotes are increasing at the expense of homozygotes
2. Allele frequency changes in one generation!
3. Equilibrium values: $p^\circ = 0.4175 \quad Q^\circ = 0.5361 \quad R^\circ = 0.0464$ are dependent on initial conditions. Its a polymorphic equilibrium.

DM can cause evolutionary change event at single locus level

It leads to excess of heterozygotes & as a result maintains polymorphisms in population.

DM is weaker at maintaining LD than AM because more heterozygotes means faster dissipation of LD.

~23 mins → Rarer phenotypes have mating advantage. Loci with DM mating system have polymorphic alleles

(?)

Big Concept

Different genetic elements in populations can display different mating systems. There's a system of mating for a specific genetic architecture within the population.

* Mating system is loci specific *

Biological implications

- Inbreeding increases homozygosity - brings together recessive deleterious alleles, which can have negative fitness consequences.
- Inbreeding depression (can also be caused by AM or genetic drift)
  - Reduced survival & fertility of offspring of related individuals.

Inbreeding depression in —

* Humans: Child mortality is higher in children of first cousins than of unrelated parents.
* Great tits - positive association b/w number of unhatched offspring and degree of inbreeding

Insights from plants

1. Inbreeding effects are more prominent when plants undergo some sort of environmental stress.

Eg: Sabatia angularis : 3 treatments - self, near outcross, far outcross (Rose pinks)

Plants showed inbreeding depression - the degree of which increased when grown in field > garden > greenhouse

2. Inbreeding effects are more likely to show up later in life cycle

Eg: In waterleaf (biennial plant)

Inbreeding dep : $\delta = 1 - \frac{w_s}{w_o}$ w_s : selfed individual fitness, w_o : outcrossed ind. fitness

δ is greater in second year than in first year

Reason : Maternal effects - through provisioning of seeds - can mask the influence of deleterious recessives until later in life cycle.

But maternal effects for a month ??

(22) 3.

Inbreeding depression can vary among family lineages

Eg: Two annual populations of common Yellow Monkeyflower

Some families showed ID, others showed no discernible effect, while others showed improved performance under ↑ inbreeding.

This shouldn't be a surprise, because F is calculated for individuals where family history becomes important.

No wonder F is variable.

Variation - at the end

1/3/22

Lecture

Strategy of genetic variation

- Local sequence change
  - Replication infedility
  - Mutagens
- DNA rearrangement
  - Recombinational reshuffling
- DNA acquisition
  - Horizonatal gene transfer

Isolation
- reproductive
- geographic
↑
Genetic diversity

Limitations of genetic variation

- Loss of DNA etc.
- Natural selection
- Genetic drift.

What does it mean when we say mutations are random?

This is random wrt fitness - not all mutations are equally probable

For a time it was thought that directed mutation as an alternative to natural selection

Luria-Delbruck's experiment on bacteria-bacteriophage system

When bacteria are mixed with phages, a fraction of them survive and become resistant.

Statistical note: If events occur with known constant mean rate and independently of the time since last event, then the probability of given no. of events occurring in a fixed interval of time or space follows Poisson distribution

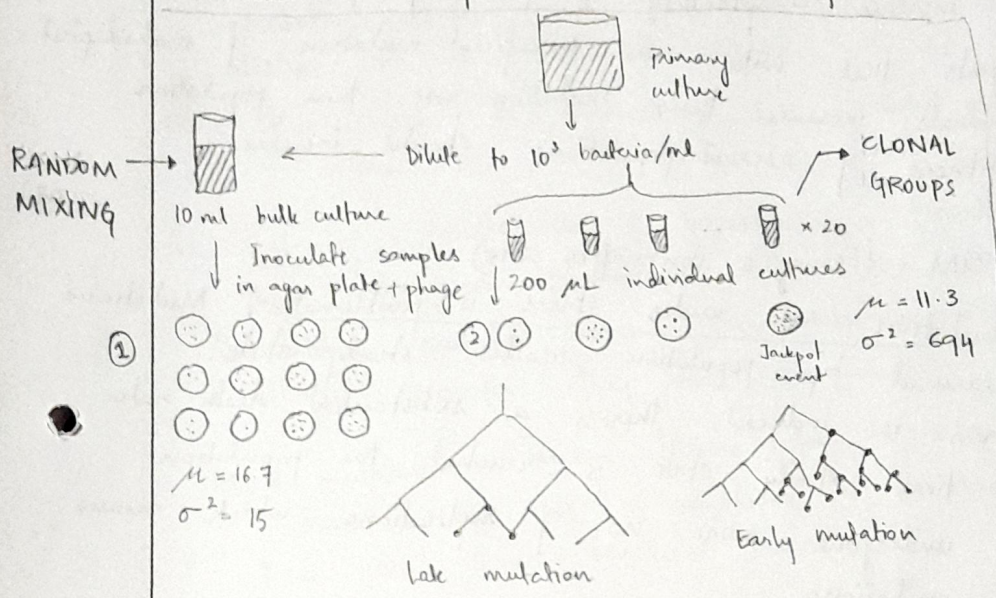
Adaptation : phenotype is induced by exposure to phage
Mutation : phenotype occurs spontaneously, prior to exposure

This can happen in two ways –

1. Adaptation hypothesis : each resistant mutation occurs as a separate random event. No clones of resistants before attack.
So, we should get Poisson - distribution of survivors.

2. Mutation : Resistant bacteria are already present in the colony before the attack. We'd get a non-Poisson results.

	Tested after Random mixing	Tested in clonal groups	No. of
Adaptation	Poisson	Poisson	
Mutation	Poisson	Non-poisson	



If mutation was being induced by the presence of phages, then we don't expect to see any difference b/w ① & ② – both Poisson

If mutation hypothesis is true, then each vial would have different no. of resistant bacteria

Are directed mutations possible ? Cairns & Foster

- They plated Lac- bacteria with lactose as the only food source
An excess of Lac+ mutations occured but not non-adaptive mutations after sometime, like directed mutation
- Later studies showed that non-adaptive mutations occured in episomes, not chromosomes of bacteria

Adaptive mutation – process that produce mutations that relieve the selective pressure, whether or not non-selected mutations occur

Mechanism of Stress Induced Mutagenesis (SIM)

In SIM adaptive or a by product of breakdown of metabolism, happens in both prokaryotes & eukaryotes (mechanism unknown).

In prokaryotes —

1. Direct increase: increased stress leads to production of toxic molecules which directly damage DNA or inhibit enzymes involved in maintaining fidelity
2. Indirect increase: stress changes expression of genes, which elevates rate of mutation by —
   - inducing expression of error-prone DNA polymerase
   - repressing genes involved in DNA repair

Is SIM important?

- SIM can increase adaptability by providing greater variation
- Theory suggests that with rare beneficial mutation, if maladapted individuals increase their mutation rate, then population mean fitness of asexual populations should increase

Challenging SIM (Frenoy & Bonhoeffer 2018)

- Measuring mutation rate under stress is problematic. Mutations are measured by population reaches steady state.
- When stress is induced, there's a substantial death rate. By the time steady state is reached, the population would've undergone more no. of replications, which means more mutations.
- Ignoring deaths leads to a systematic overestimation of mutation rates under stress, by about 50%.

Measurement of mutation rates

- per genome per generation (U) — more imp. from evolutionary pov since fitness is integrated over entire genome
- per site (bp) per generation
- per gene per generation

Before sequencing, there were major methodological & statistical challenges to measuring mutation. Its better with cheap sequencing

Insights about mutation : Qualitative

* Transitions (A↔G, C↔T) are more common & more drastic than transversions
* Mutation rate depends on no. of germline cell divisions per generation
  → Accumulation of mutations in older individuals
  More mutations in males owing to more pre-meiotic cell division
* Mechanisms well known
* Different polymerases lead to diffirent mutation patterns and their function also causes mutagenic effects of methylation & mutational hotspots.

Deleterious mutation rate is likely to be U > 1 for eukaryotic species.

- Fidelity of replication is highly dependent on its local sequence context ie. adjacent base pairs.
- Different neighbouring nucleotides can alter site-specific mutation rates by 75-times
- Sites neighboring GC pairs have higher mutation rate

Quantitative insights

Tremendous variation in mutation rates across various taxa when measured as per bp per generation

Given most mutations are deleterious*, why don't mutation rates evolve to zero?

When scaled by genome size, mutation rate (per genome per gen) is surprisingly similar across organisms.

Drake's rule : mutation rate per bp is inversely proportional to genome size so that no. of mutations per genome per generation is roughly constant.

Inference: This strongly implies that this const. rate is highly evolved and must have been shaped in response to evolutionary forces of a very general nature, forces independent of kingdom & niche

Lecture - Rate of Evolution

2/3/22

There are some issues (Lynch 2010)

In non-eukaryotes, there's a negative relationship b/w mutation rate and genome size. In eukaryotes, there's a * nice positive relationship.

Patterns of mutation rates

Hypothesis 1

1. Direct evolution of mutation-rate modifiers

There are some sites in bacteria that drive mutation when they are modified

Support : Lenski's 12 LTE lines

3 of these evolved 100 times increased mutation rate

Later investigations suggested that this was because of mutative alleles were hitchhiking (linked) with beneficial alleles. This wasn't direct selection

Q1 : This works in asexual populations. What happens when there is recombination?

Q2: Cost of maintainance of hypermutators (simulation based)

They simulated populations with 10x, 100x & 1000x mut. rates

Intermediate mutation rates - these mutators become fixed, but after sometime, they will disappear

1000x mutators - mean fitness goes to 0.99, but population never rises. Cost of mutations is too high, even if there are some beneficial mutation

Muller's Ratchet - process by which, in the absence of recombination genomes will irreversibly acquire deleterious mutations

In an asexual population, once the fittest individual has at least one deleterious mutation, there is no way by which the fittest genotype will come back. The population accumulates deleterious mutations & becomes extinct - Mutational meltdown

Evidence : Sprouffske 2018

Engineered identical strains of E.coli with different mutation rates (S, M, L, XL) s.t. XL = 1395. Grown for 3000 generations and exposed to 90 different environments. And did whole genome population sequencing.

Results

1) Higher mut. rates (mR) lead to greater genetic diversity, but this is beneficial for intermediate mutators, where populations adapted faster and thrived better

2) Highest mutators showed reduced adaptation & didn't thrive in any of 90 environments. They also experienced a dramatic decrease in mutation rate.

3) However, there's no simple correlation b/w ancestral mutation rate and stress tolerance after evolution.

Hypothesis 2

2. Existence of a cost of fidelity associated with replication speed
   ⇒ Fastest replicators should have highest mutation rate
   But in reality, the opposite is observed.
   ↳ too variable, hard to establish correlation

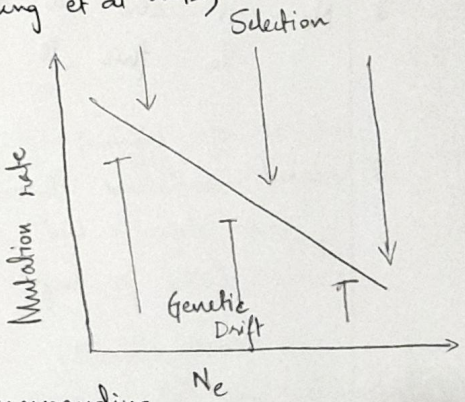
Hypothesis 3

3. Physicochemical limits to the accuracy of replication and repair processes.
   It should act similarly in many organisms. This doesn't explain the variability of mutation rates observed in similar physiologies.

Hypothesis 4

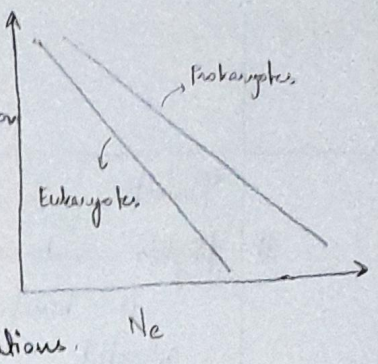
Drift barrier hypothesis (Lynch 2010, Sung et al 2012)

- Selection tries to reduce mutation rate
  The effectiveness of selection is reduced by the amount of drift faced by the population.
- Effectiveness of selection is proportional to effective population size (Ne)



Ne: Ne of a population is the corresponding size of ↑↓ population which is undergoing the same drift as the population in question.

There is indeed a negative relation between mutation rate and N_e



But, selection is less effective at reducing per base mutation rates in prokaryotes than in eukaryotes, despite the fact that such selection is expected to be stronger in non-recombining populations.

Explanation for difference in intercepts

Prokaryotic genomes : <5000 protein-coding genes

Eukaryotic genomes : 7000 - 30,000 genes

Its possible that magnitude of selection is not simply a function of per-site mutation rate but also of the genome wide deleterious mutation rate. Since prokaryotes have smaller genomes, they provide smaller target for origin of such mutations than eukaryotic genomes.

⇒ Magnitude of selection (acting to reduce /bp mut. rate) is expected to scale positively with effective size of a genome

As expected, effective genome-wide mutation rate is inversely proportional to N_e, independent of taxa

* Drift-barrier hypothesis is a general one that can also be applied to other phenotypes and the selection pressure on them.

* N_e is also a function of the ecology of the species. So, this is a truly eco-evolutionary theory.

6/3/22

Variation - Recorded

Why is variation important?

- Variation is the raw material for all microevolutionary processes, in particularly selection
- Novel variation might be needed for formation of different groups (macroevolutionary changes)

If we accept that macroevolution happens at fast time scales, through ~~macroevolutionary~~ macromutations, it this is a departure from the canonical view.

Where is the variation?

↳ At every level of biological organisation. Different ways of looking at it.

Eg. Skin colour variation in humans — where does variation come from?

↳ Variation in skin colour at birth

→ Due to exposure to sun

→ Variation in how individuals differ in responding to ~~variation~~ radiation.

- Genetic variation : variation due to differences in genetic code
- Environmental variation : Arises when external factors influence how much protien is made or how they work
- GxE variation: Genotype-by-Environment interaction is the result of how different genetic components react differently to environmental influences.

→ Genetic variation

We know a lot about this. This is thought to be ~~be~~ the primary source evolutionarily relevant variation because its heritable. Still considered important, but people are trying to understand other sources (goal of EES).

* Eg. Stochastic variation in gene expression. (Elowitz 2002)

Built strains of E.coli incorporated cyan (cfp) and yellow (yfp) with identical promoter & equidistant from origin of replication on either side. This unearthed a significant amount of noise in gene expression.

(20)

Environmental variation

Eg: • Inducible defenses in Daphnia juveniles that smell phantom midge larvae (predator) grow neck teeth & other defenses

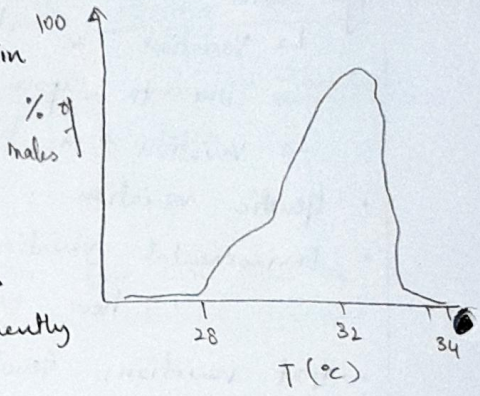
• Discrete seasonal polyphenism in Nemoria sp. caterpillars.
Summer brood feeds on leaves & resembles oak twig.
Spring brood feeds on inflorescence (catkin) and they resemble catkins.

- Classically, this was not important since its not heritable. But now we know, many environmentally-induced variation can pass across generations (epigenetics & plasticity)

Big concept
Selection deals not with genotype, but with its dynamic properties, its reaction norm (i.e. the phenotype which is a result of GxE interaction), which is the sole criterion of fitness in the struggle for existence - Dobzhansky.

→ Genotype x Environment variation

★ Eg: Temp-dependent sex determination in leopard gecko



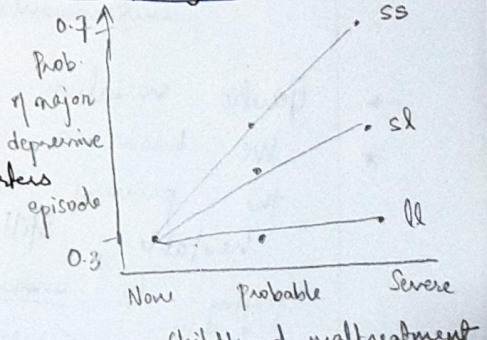
Intermediate T is more likely to produce males.

This response varies across families
ie different genotypes respond differently to environment.

★ People with different genotypes differ in sensitivity to maltreatment during childhood.

3 genotypes of serotonin transporter gene
It removes serotonin from brain
Individuals with ss encode less transporters than those with ll.

ss are more sensitive to trauma in childhood

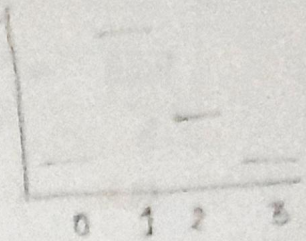


Can this sensitivity itself be selected?

* **Evolution of plasticity in tobacco hornworms** (Suzuki & Nijhout 2006)

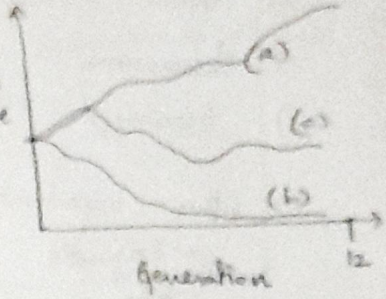
Typical caterpillar is green in colour, but there's also a black strain. When black mutants are exposed to heat shock, some of them turn green to varying degree

Colour score : 0 → not sensitive
3 → very sensitive (v green)

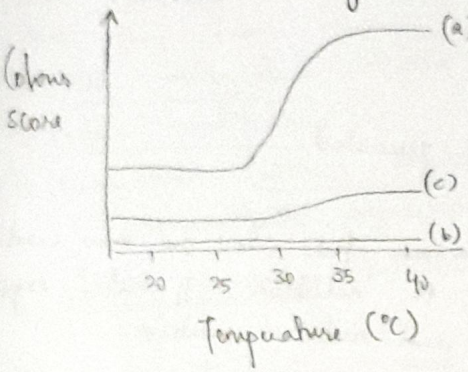


There Three selection lines were established :
- Selection for high plasticity (a)
- Selection for low plasticity (b)
- control (c)

Average colour of caterpillar in each line after giving heat shock



They also reared these caterpillar at various temperature and heat shocked them & measured avg colour score



Variation in genotype to respond to environmental cues can be selected for!

low plasticity : no variation

Populations that live in <u>variable & unpredictable environments</u> can <u>benefit from higher levels of phenotypic plasticity</u> (PP).

Populations that live in constant environments don't need PP, but they still have it.

Does PP aids evolution or happens it?

(23)

Classifying variation based on effect magnitude.

a) Continuous / small effect variation – small difference in expression level
b) Discrete / large effect variation – can lead to new functions or even reproductive isolation

Hard to apply this practically.

Based on how it is generated

It can be studied at the level of –

i) Alleles – mutations, transposons
ii) Genes – gene duplication, retropositon, unequal crossing over
c) Chromosome – chromosomal translocation
d) Genome – changes in ploidy

Other important stuff –

- Epigenetic
- HGT
- Behaviour / culture
- Contextually neutral variation
- Evolutionary capacitance

7/8/22

Lecture – Variation II

Ways in which variation can be generated.

1. Recombination

Very important for sexual organisms. Puts loci in new combinations and exposes recessive allele to selection. It also exposes deleterious alleles and hence can reduce variation.

2. Environmentally induced variation – Waddington expt (1953)

Drosophila normally have two crossveins. When pupae (21-25 hrs) are subjected to heat (104 F for 2 hrs), there crossveins in adults are broken.

There two phenotypes were subjected to selection – crossveinless phenotype went from ~60% to ~100%.

After 20 generations, the pupae were crossveinless even without the heat shock. Acquired variation had become heritable and fixed in the population.

Genetic assimilation: Process by which environmentally induced population variation responds to selection and becomes stably inherited.

Different way of looking at how evolution may work (West-Eberhard)

1. Trait origin : mutation or env. change causes appearance of a developmental variant expressing a novel trait
2. Phenotypic accommodation: to the new trait made possible by inherent pre-existing plasticity of developmental system
3. Initial spread of new variant facilitated by its recurrence in the population, if the initial change is environmental (a rare mutation would take a long time to spread).
4. Genetic accommodation : fixation by allelic substitution as a result of standard selection. Now even in environment changes back, the phenotype won't change.

This is a phenotype-first view

→ Epigenetic variation

Anything that leads to change in phenotypic level without altering DNA sequence

How they work –

1) DNA methylation – gene packing & tx rate changes
2) Histone modification
3) micro-RNA (miRNA) regulation - 17-25 bp long RNA that are epigenetically regulated.

Eg: Licking grooming behaviour in rat mothers

- High nurtured pups become relaxed, high nurturing females themselves and vice-versa
- At birth, promoter of Glucocorticoid Receptor (GR) protein is highly methylated. Methylation decreases when licking-grooming happens & GR is expressed in hippocampus. This makes the rat/mouse more relaxed because uptake of cortisol is increased
- How stable is the inheritence? Not much is known about this specific case, but transgenerational epigenetic inheritence is important. Examples –

1) When pregnant female is treated with vinclozolin, it alters methylation patterns. The rat pup is born with several abnormalities which continues down to all male progenies.

2) Linaria vulgaris – comes in 2 types – bilateral and peloric form
They were thought to be 2 alleles of a gene
The lcyc gene is excessively methylated in peloric mutant & is stably inherited. The gene sequence itself is identical
When methylation level is varied experimentally, extra spurs are developed on bilateral flowers

3) Methylation status of GR gene in mothers and their children
10-19 years after birth. They also recorded maternal exposure to intimate partner violence (IPV) during & after birth
Mother's methylation of mother's GR gene was not affected by IPV. IPV before or after pregnancy didn't affect child's methylation rate whereas when there was IPV during pregnancy, methylation level of GR gene of those children during adolescence was high

4) Adult offspring of Holocaust survivors had greater risk for development of PTSD, depression & anxiety disorder. This was true even though they were born after WWII.
Survivors & offspring had undergone epigenetic modifications for FKBP5 gene, a regulator of GR.

Epigenetic variation in quantitative traits – Cajito 2014

Epialleles : alleles with same DNA sequence but different methylation patterns

They created plant populations of Arabidopsis that contained large variation in epialleles for 2 traits: flowering time and primary root length

They were found to be stably inherited over 8 generations.
Also found that ~30% of differentially methylated regions induced in expt were also found in natural populations

→ Cryptic genetic variation (Rutherford & Lindquist 1998)

When drosophila were given heat shock, it was observed that a lot of new phenotypes were being developed. A lot of variation would express at phenotype level if they weren't suppressed/masked by chaperone genes. They are making the existing variation cryptic.

The idea was that when heat shock is given, the hsp genes become busy dealing with heat shock, which allows all these phenotypes to be expressed

In arabidopsis, they treated the plant with GDA which allowed expression of variation in morphology

Cryptic variation can be (potentially) adaptive - Rohner et al 2013

Rohner et al studied cave fishes where they adapt to low light by losing eye because its expensive to maintain

Another thing that changes is salinity - ~~underwater~~ caves water has low salinity.

When fish are put in low salinity condition, variation in eye size and orbit size increases. Mean hardly changes

Similar results can be induced by radicicol, a drug which results ~~eye size & orbit size~~ in increased variation

They treated fish with radicicol, chose 2 fish from F1 with small eye size to get F2. F2 individuals have significantly smaller mean eye size & orbit size.

The cryptic genetic variation that's exposed through radicicol treatment is selectable.

Evolutionary capacitance

The phenomenon by which phenotypic variation can be exposed to selection in response to a cue, typically environmental. but could be internal also.

Eg. In yeasts, Sup35 protien helps terminate translation when stop codon is reached. In [PSI+] conformation, Sup35 aggregates, becomes less soluble & conc. of free protien decreases So, stop codon becomes less effective & the region after it is also translated. This can cause significant ~~variation~~ phenotypic change based on nature of the variation. Under normal conditions, [PSI+] is tolerated, but when stressed, [PSI+] will exhibit more variation.

Developmental systems are often robust to environmental perturbations, when the robustness capacity is overwhelmed, they end up throwing out phenotypic variation

This in some sense is a hinderance to adaptive evolution. But evolutionary capacitance is a specific type of variation — 'just in time' release of variation, which can actually facilitate evolution. If there's a stable environment, then most variation would be winnowed down and when environment changes, it's possible that there isn't enough variation for selection to act on.

Lecture - Variation III

Major insight on variation

→ Life is fundamentally conserved — genetic code, transcription, translation etc.

This suggests that either there are very few ways of making life work, ~~or~~ or historical contingency (frozen accident), or both.

Since life is so constrained, stabilizing selection could be the dominant mode of natural selection. Comparative genomic data suggests that nucleotide substitutions are often synonymous (coded aa is not changed) than non-synonymous.

→ Genome of complex organisms contains a lot of 'junk' DNA

1-5% of human genome codes for proteins. Out of that, only 5% appears to be constrained. This encodes RNAs which regulatory role & other functions we don't know about.

We don't know yet doesn't mean it doesn't have a function.

A possible role of these sequences might be in the origin of new genes which might be the missing link to macroevolution.

→ Gene number doesn't predict organismal complexity

Human : 20,000 E.coli : 4,400

Arabidopsis : 27,500 Rice : 41,000

is evolutionary capacitance observed in a variety of organisms? Can it be a selection for evolutionary capacitance?

Possible resolutions -

1) Alternative splicing & other mechanisms increases the no. of different proteins in an organism
2) Complexity might depend more on molecular interactions than no. of genes or proteins.
3) Gene expression regulation & post-translational modification can increase organismal complexity.

→ Horizontal gene transfers are prevalent

Its pretty common in prokaryotes, less so in eukaryotes.

Eg Sea slug : Elysia chlorotica

Juveniles feed on alga, Vulcheria. Its not digested, instead it retains plastids and use it to synthesize carbs.

~52 genes of algae have become stably incorporated in the slug, which supplies proteins required for plastids.

Recent studies have cast doubts on HGT. Algal plastids can survive without nuclear genes.

Three ways in which HGT happens -

- Transformation: cells take up free DNA
- Conjugation: living cell transfers genetic material to another cell
- Transduction: DNA transfer through virus.

- Practical problem: antibiotic resistance
- Conceptual problem: Shift from binary tree to reticulate tree

Its a problem for phylogeny reconstructions, doesn't take away from majority of genes are vertically transmitted

Soreh (2007) studied attempted movement of 2,46,045 genes from 79 prokaryotic genomes into E.coli and identified genes that consistently failed to transfer. They found that genes related to translation are most recalcitrant to HGT — so these genes could be used for reconstruction.

But much more data is needed.

Ku et al 2015 - also ~~tried~~ sequenced a lot of proteins across prokaryotes and eukaryotes.

Ku et al showed that -

1) gene transfer from prokarya to eukaryotes is episodic and coincides with major evolutionary changes
2) gene inheritance in eukaryotes is vertical
3) continuous HGT doesn't contribute to long term gene evolution in eukaryotes.

→ Gene duplication is primary source of new genes
Every genome has some duplicated genes. One copy retains old function whereas the other copy acquires new function.
Recently, it's been proposed that ~~non-sequencing~~ non-coding sequences are the source of new genes

→ Changes in protein fn & gene expression are both important to phenotypic evolution.
Conjecture: Morphological evolution prolly through gene expression changes
* Physiological evolution through protein function changes.

→ Interspecific genetic polymorphisms are abundant & largely neutral
In humans, nucleotide diversity is 0.1% ⇒ any two alleles of some gene differ by 1 out of 1000 bases.
This neutral variation was considered unimportant for long time
Eg: codon bias exists. So neutral variation can cause fitness change
If this neutrality is conditional, this means that this is an enormous amount of source/raw material for evolution.

Tapes discussion

9b) Its disassortative mating because observed freq of heterozygotes is more than expected freq of heterozygotes

10) Given genotypic frequencies of 3 loci.
From this, calculate gene frequencies — $p = P + \frac{1}{2}Q$
$q = R + \frac{1}{2}Q$

From p and q, calculate expected heterozygote frequencies by $Q' = 2pq$

The calculate $f = 1 - \frac{Q}{Q'} = 0.1$

If we get $f = 0.1$ for 3 randomly chosen loci, that means either these loci are linked and there's AM, or this system is inbreeding.

11) The system is not in HWE for 2 locus system because $f(AAbb) = 0$

At equilibrium, recombination doesn't matter. So genotypic freq are given by multiplication of gene frequencies
i.e $f(AABB) = p^2q^2$

11.c) when s changes, if system is in equilibrium, if doesn't affect anything. So genotype frequencies won't change.

12.b) Chances of getting Rh+ child when both parents are Rh+

	DD	Dd	dd
DD × DD	P^2	1	
DD × Dd	2PQ	0.5	0.5
Dd × Dd		0.25	0.5	0.25

$P' = 0.36$

$Q' = 0.288$

$R' = 0.057$

Fraction : $\frac{P' + Q'}{P' + Q' + R'} = 91.8\%$

12b → got the right answer

8 → ½ mark lost for writing $1 - \frac{1}{n}$

Lecture 23/8

Genetic Drift

HWE assumes an infinitely large population size What if this is not true?

GD - change in allele frequencies due to random sampling
Random sampling irrespective of their effects.
It's the effect of stochasticity on a finite population

Figulus simulation

When population is infinitely larges, the frequency of alleles in gametes will be same as that of the population
This need not be true for small populations.

Buri's Experiments (1956)

- founded 107 populations with 8 pairs each drosophila
- All founders were heterozygous - bw^+/bw^-, all genotypes were distinguishable & no known fitness effect
- Each generation, he counted no. of genotypes, & collected 8 random pairs for next generation. He did this for 107 & 19 generations.
- Results
  - All start out with $f = 0.5$, but by 19th gen, bw^{75} had been lost from 30 populations and 100% fixed in 28 populations.

GD is the random change in allele frequency due to sampling error in a finite population.
It is a microevolutionary force.
Natural selection can explain adaptation, whereas Drift can't explain it.

Simulation: If we increase population size, the effect of drift decreases and more of the populations get fixed.

- Allele frequencies change randomly, no directionality ⇒ can't explain adaptation
- No tendency to return to ancestral allele freq. departure accumulates over time.
- Only 2 fixed alleles states for an allele. GD leads to loss of genetic variability in a population.
  This also increases avg probability of identity by descent (inbreeding coefficient) in a population.
- GD can increase differences in allele frequency b/w finite subpopulation.

Bottle neck effect
Change in genetic variation that can occur when population size is sharply reduced, followed by recovery in numbers.

Founder effect
Change in genetic variation when a new population is established by a smaller no. of individuals from a large population

Mechanistically, these two are the same, conceptually slightly diff.

Drift is a continuously occuring process. These effects are discrete, special-case events that enhance the effects of drift.

Consequences: Humans in India
Due to within castes marriages, genetic diversity of several communities are relatively low. Some communities have had founder effect.
Some diseases are prevalent in some communities

Eg 1. Parsis

Parsis migrated from Iran to West coast of India, over centuries in male-dominated migrations & took local wives.

So, their Y-chromosome maps back to Iranian lineages, while mt-DNA maps to south-east asians.

Genetic studies show ~~low~~ high genetic homogeneity & high rates of breast cancer, bladder cancer etc.

Also high rate of inbreeding - Bombay blood group.

Eg. 2 Vysya community (South India)

Now 2.2 crores, however they've had strong founder effect.

They have 100-fold higher rate of butyrylcholinesterase deficiency which makes muscle relaxants (anasthaetics) lethal to them.

Genetic Drift & Linkage Disequilibrium (LD)

Drift contributes greatly to LD, with increasing no. of loci.

As no. of loci increases, ↑ gametic genotypic combination ⇒ lesser no. of individuals per genotype ⇒ greater effect of drift.

Eg 1 Colour blindness & Anaemia in Italy

↳ X-linked X chr also contains -

- Glucose-6-phosphate dehydrogenase gene mutation, which leads to haemolytic anaemia
- 2 tightly linked genes for colour vision.

In Italy, one population west of Appenine mountains - only one G6PD (Med1) population + Colour blindness

In Sardinian population, (founder effect) 90% of G6PD has Med2 population and no CB

Both populations are showing LD, but in opposite directions.

In South Italy, CB can be used as marker for G6PD deficiency & therefore haemolytic anaemia.
But this is not true for the Sardinian population.

Therefore: Disease/marker association studies should not be * generalized beyond the actual populations studied *

Different populations have different gene-marker associations, often due to historic reasons.

In US, African Americans have greater incidence of hypertension & high frequency of R_0 allele at Rh blood group.
Americans of English ancestry have low incidence of hypertension and very low freq. of R_0.

If you pool two populations & do gene-disease association study, then we might end up with false positive correlations.
Hence, appropriate populations must be carefully selected.

Genetic drift & Evolutionary outcomes.

Suppose 100% AM for phenotype determined by two loci — we saw that multiple outcomes possible based on initial freq.
Once we bring drift into the picture, it's hard to predict equilibrium/final outcomes, even if we know initial freqs.

29/3/22

Lecture

Effects of Genetic Drift

Effect based on —

1. Rate of change in F inbreeding coeff.
2. Variance of allele frequencies
3. Any other genetic feature of population (eg. time to fixation)

What is the size of the ideal population that undergoes the same amount of drift as the given population?

Different populations have different stochastic processes. So, we need a reference population, or ideal population.

Effective population size (N_e)

Defn: No. of individuals in an ideal population that has a value of any given population genetic quantity that is equal to the value of that quantity in the population of interest.

Important points —

for the same population,

1. There can be different N_e s based on which parameter is being tracked
2. These N_e s can be close values or maybe not
3. Same measure of N_e might have different values based of on different reference points.

All of this should be qualified.

Ideal population

Consider a closed population of hermaphroditic diploids in which selfing is allowed. In each generation, following processes take place —

1. Each individuals produce vast no. of sperms & eggs ie. no selection
2. Individuals all spawn into a common gamete pool
3. Gametes are thoroughly mixed, and zygotes are formed through random mating.
4. There's density dependent mortality, not based on genotype which reduces zygotes to N surviving adults.
   No age structure, discrete populations.
5. No change in population size.

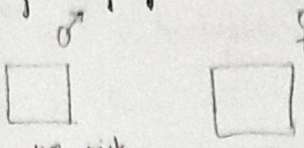
Effective inbreeding coefficient (N_{eI})

At $t = 0$: N, $\bar{F}_0 = 0$ (no related individuals)

$t = 1$: N, $\bar{F}$ is affected by selfing & existing

Probability of selfing :

A_1A_2 = the alleles are not related by IBD



Suppose we pick a sperm by individual x with genotype A_1A_2

Prob. of picking ovum by x - $\frac{1}{N} \times \frac{1}{2}$

$\frac{1}{2}$ - prob. that ovum has A_1 when sperm we've picked also has A_1 allele

Prob. of selfing $= \frac{1}{2N}$

Prob. of not selfing $= 1 - \frac{1}{2N}$

Prob of IBD not by selfing $= \left(1 - \frac{1}{2N}\right) \times \bar{F}_0$

$$\Rightarrow \bar{F}_1 = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\bar{F}_0 = \frac{1}{2N}$$

$t = 2$: N, $\bar{F}_2 = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\frac{1}{2N}$

At this point, it becomes easier if we switch to heterozygosity of population defined by -

$$H_t = 1 - \bar{F}_t$$

where $F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\bar{F}_{t-1} = 1 - H_t$

$$\Rightarrow H_t = 1 - \frac{1}{2N} - \left(1 - \frac{1}{2N}\right)\bar{F}_{t-1}$$

$$H_t = \left(1 - \frac{1}{2N}\right)\left(1 - F_{t-1}\right) = \left(1 - \frac{1}{2N}\right)H_{t-1}$$

Based on this, $\bar{F}_0 = 0 \Rightarrow H_0 = 1$

$$\Rightarrow H_1 = \left(1 - \frac{1}{2N}\right)H_0 = \left(1 - \frac{1}{2N}\right) \Rightarrow H_t = \left(1 - \frac{1}{2N}\right)^t$$

$$H_t = 1 - \bar{F}_t = \left(1 - \frac{1}{2N}\right)^t$$

$$\Rightarrow \bar{F}_t = 1 - \left(1 - \frac{1}{2N}\right)^t \quad \Rightarrow \quad \frac{1}{2N} = \left(1 - \bar{F}_t\right)^{1/t}$$

$$\Rightarrow \boxed{N_{ef} = \frac{1}{2\left[1 - (1 - \bar{F}_t)^{1/t}\right]}}$$

N is the population size required to get the value $\bar{F}_t$ at time t.

$\bar{F}_t$ is calculated from the reference population

As long as $\bar{F}$ can be calculated empirically, N_{ef} subsumes all sources of drift by comparing in its effect to corresponding amt. experienced by an ideal population

N_{ef} depends on many factors, but not census size.
N_{ef} can be $<$, $>$, $=$ to N.

Eg: A planned avoidance of mating b/w relatives can take $N_{ef} >$ census size

Variance Effective Population size (N_{ev}) Templeton ?

Defn: N_{ev} measures how rapidly allele frequencies are likely to change and/or how rapidly isolated subpopulations diverge from one-another under genetic drift

Formulae: σ_t^2 : variance of allele freq after t generations

$$\sigma_t^2 = pq\left[1 - \left(1 - \frac{1}{2N_{ev}}\right)^t\right] \Rightarrow \boxed{N_{ev} = \frac{1}{2\left[1 - \left(1 - \frac{\sigma_t^2}{pq}\right)^{1/t}\right]}}$$

p, q : allele frequencies.

When N fluctuates from generation to generation —

$$N_{ef} = \frac{t}{\frac{1}{N(0)} + \frac{1}{N(1)} \cdots \frac{1}{N(t-1)}}$$

$$N_{ev} = \frac{t}{\frac{1}{N(1)} + \frac{1}{N(2)} \cdots \frac{1}{N(t)}}$$

N_e ~ Harmonic mean of pop. size.

$N_{ef} = N_{ev}$ if $N(t)$ is constant for all t. This is the only case when they're equal (but not always true)

Here are several equations for N_{ev} vs N_{ef} under a variety of deviations. — For example —

Similar but not the same

Empirical example : N_e of wild rhino population in Africa

- N_e varies greatly, it need not be less than N.
- Estimates of N_{ef} & N_{ev} are not the same — it's critical to state which measure of N_e is being considered

50/500 Rule in conservation biology

Franklin 1980 : focus on species whose N_e is >50 in short term and <500 in long term.

Templeton 2020 : Stated differently, it should start with more than 50 individuals and go to 500 as quickly as possible

Both parts of the rule are empirically unsupported

It's possible to decrease/increase N_e by avoiding in-breeding

The latter value depends on many factors — mutation rates, mating etc. So this rule doesn't make sense

How to measure N_e ?

Measuring $\bar{F}/p$ is not easy. One method — estimate $\bar{F}_0$ and $\bar{F}_t$ from genetic data —

$$\Delta F = \bar{F}_t - \bar{F}_0$$

① $$N_{ef} = \frac{1}{2\left[1 - \left(\frac{1-\bar{F}(t)}{1-\bar{F}(0)}\right)^{1/t}\right]}$$

② $$N_{ef} \simeq \frac{t}{2\left[\bar{F}(t) - \bar{F}(0)\right]}$$

③ $$N_{ef} \simeq \frac{t}{2\Delta\bar{F}}$$

when $\bar{F}_0$ are small

for $$N_{ev} = \frac{4 N_m N_f}{N_m + N_f}$$

, Also the harmonic mean of N_{ev} of every timestep.

Another may measure N_{ev} in each generation using demographic data & then get an overall estimate using HM —

$$N_{ev} = \frac{t}{\frac{1}{N(1)} + \frac{1}{N(2)} + \cdots \frac{1}{N(t)}}$$

Ultimately : Which method to choose depends on which N_e is needed for what purpose

Interpret N_es carefully!

Lecture

5/4/22

Selection

1-locus - 2-allele haploid case

Say 2 alleles - A_1 A_2

Allele frequency: p $1-p$

2 types of selection modelling - Viability selection
Fertility selection

$$\text{Adults}^{(t)} \xrightarrow{\text{Fertility selection}} \text{Offspring} \xrightarrow{\text{Viability selection}} \text{Adults}^{(t+1)}$$

As long as differential fertility & viability are both constants -

$$\boxed{N(t+1) = F_1 \cdot V_1 \cdot N(t)}$$

In viability selection, both fertility and survivorship can be considered because what is important is the constant in $N(t+1) = \underline{k}\, N(t)$

	A_1	A_2
F_1	p	$(1-p)$
surviv/fertility	w_1	w_2
	pw_1	$(1-p)w_2$
F_2	$\frac{pw_1}{pw_1 + (1-p)w_2}$	$\frac{(1-p)w_2}{pw_1 + (1-p)w_2}$
	$p' = \frac{pw_1}{\bar{w}}$	$\frac{(1-p)w_2}{\bar{w}} = 1-p'$

Scaled so that if sums to 1

$$\Delta p = p' - p = \frac{pw_1}{\bar{w}} - p = \frac{p}{\bar{w}}(w_1 - \bar{w})$$

$$\Delta p = \frac{p}{\bar{w}}(w_1 - pw_1 - w_2 + pw_2)$$

$$\Delta p = \frac{p}{\bar{w}}(w_1 - w_2)(1-p)$$

$$\therefore \boxed{\Delta p = \frac{pq(w_1 - w_2)}{\bar{w}}}$$

If there are two arrenals / microbes with differential fitness, then one of them will outcompete the other.

1-locus - 2 allele diploid case

	A_1	A_2	
gene freq	p	q	
	A_1A_1	A_1A_2	A_2A_2
F_1	p^2	$2pq$	q^2
	w_{11}	w_{12}	w_{22}
F_2	$\frac{p^2 w_{11}}{\bar{w}}$	$\frac{2pq w_{12}}{\bar{w}}$	$\frac{q^2 w_{22}}{\bar{w}}$

where $\bar{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22}$

Unlike haploid case, here fitness is defined for genotype, not the allele. How to define allele fitness?

$$p' = \frac{p^2 w_{11}}{\bar{w}} + \frac{pq w_{12}}{\bar{w}}$$

$$p' = \frac{p}{\bar{w}}\left(p w_{11} + q w_{12}\right) = \boxed{\frac{p w_1^*}{\bar{w}}}$$

Marginal allelic fitness

Is the average fitness of all individuals who carry at least one copy of the allele

A_1A_1	A_1A_2
$p w_{11}$	$q w_{12}$

↓ conditional probability

If one of the alleles is A_1, what's prob. that other allele is also A_1? — p

$$w_1^* = p w_{11} + q w_{12}$$

$$w_2^* = p w_{12} + q w_{22}$$

$$\bar{w} = p\left(p w_{11} + q w_{12}\right) + q\left(p w_{12} + q w_{22}\right)$$

$$\therefore \boxed{\bar{w} = p w_1^* + q w_2^*}$$

$\bar{w}$ can also be written as -

$$\bar{w} = p^2 w_{11} + 2p(1-p) w_{12} + (1-p)^2 w_{22}$$

$$\bar{w} = p^2 w_{11} + 2p w_{12} - 2p^2 w_{12} + w_{22} - 2p w_{22} + p^2 w_{22}$$

$$\frac{d\bar{w}}{dp} = 2p w_{11} + 2 w_{12} - 4p w_{12} - 2 w_{22} + 2p w_{22}$$

* Frequency independent assumption

: Assuming w_{11}, w_{12} and w_{22} don't depend on p

$$\frac{d\bar{w}}{dp} = 2\left(pw_{11} + w_{12} - pw_{12} - pw_{12} - w_{22} + pw_{22}\right)$$

$$= 2\left(pw_{11} + w_{12}(1-p) - \left[pw_{12} + w_{22}(1-p)\right]\right)$$

$$= 2\left(\left[pw_{11} + qw_{12}\right] - \left[pw_{12} + qw_{22}\right]\right)$$

$$\boxed{\frac{d\bar{w}}{dp} = 2\left(w_1^* - w_2^*\right)}$$

Recall : $p' = p\frac{w_1^*}{\bar{w}}$

$$\Rightarrow \Delta p = p' - p = \frac{p}{\bar{w}}\left(w_1^* - \bar{w}\right) \qquad \bar{w} = pw_1^* + qw_2^*$$

$$\Delta p = \frac{p}{\bar{w}}\left(w_1^* - pw_1^* - qw_2^*\right)$$

$$\Delta p = \frac{pq}{\bar{w}}\left(w_1^* - w_2^*\right)$$

$$\therefore \boxed{\Delta p = \frac{1}{2}\cdot\frac{pq}{\bar{w}}\cdot\frac{d\bar{w}}{dp}}$$: Sewall Wright's equation for adaptive landscape.

Insights –

1. $\frac{pq}{\bar{w}}$ is a positive quantity $\Rightarrow$ Δp sign is dependent on $\frac{d\bar{w}}{dp}$

2. How does $\bar{w}$ (avg. fitness) change with p?

$\bar{w} = p^2w_{11} + 2pqw_{12} + q^2w_{22}$ $\Rightarrow$ $\bar{w}$ is quadratic in p

$\Rightarrow$ it has one maxima or minima

$$\frac{d\bar{w}}{dp} = 2\left(pw_{11} + w_{12} - 2pw_{12} - w_{22} + pw_{22}\right)$$

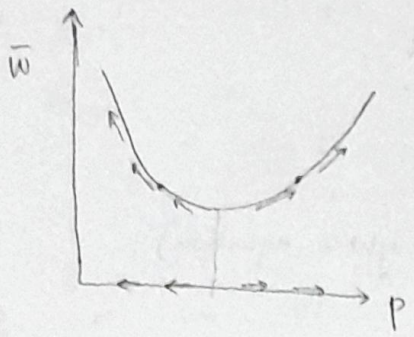
$$\frac{d^2\bar{w}}{dp^2} = 2\left(w_{11} - 2w_{12} + w_{22}\right)$$

If this quantity is –
positive : minima
negative : maxima

+ve : $w_{11}, w_{22} > w_{12}$

-ve : $w_{11}, w_{22} < w_{12}$

$w_{11}, w_{22} > w_{12}$

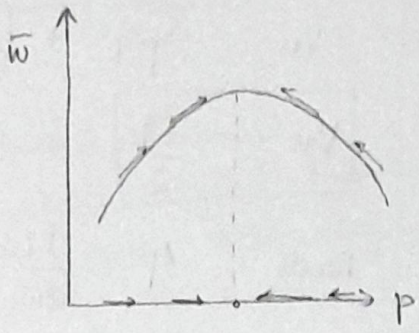


Underdominance

Heterozygote is less fit than homozygotes

Ultimately, $p = 0$ or $p = 1$ such that $\bar{w}$ increases

$w_{11}, w_{22} < w_{12}$



Overdominance

Heterozygote is fitter than both homozygotes

Some intermediate value of p is stable where $\bar{w}$ is maximum.

We can also have cases where $\bar{w}$ monotopically increases or decreases with p. But there too, average fitness of the population always increases as a result of selection.

This is taken as a given - which is not necessarily true

6|4|22

Lecture

Consider 1-locus - 2-allele case. The probability of pulling out A_1 from individual of genotype -

A_1A_1	A_1A_2	A_2A_2
1	0.5	0

$A_1 \rightarrow p$

$A_2 \rightarrow 1-p = q$

Expectation of selecting A_1 from the population -

$$E = 1 \cdot p^2 + 0.5 \times 2pq + 0 \cdot q^2$$

$$E = p^2 + pq = p(p+q)$$

$$\boxed{E = p}$$

$$\text{Var} = \Sigma (x_i - \bar{x})^2 \cdot \text{freq}$$

$$\text{Var} = \underbrace{(1-p)}_{q}^2 \cdot p^2 + \left(\tfrac{1}{2}-p\right)^2 2pq + (0-p)^2 q^2$$

$$\text{Var} = 2p^2q^2 + \left(\tfrac{1}{4} - p + p^2\right) 2pq \qquad -p+p^2 = -p(1-p) = -pq$$

$$\text{Var} = 2p^2q^2 + \left(\tfrac{1}{4} - pq\right) \cdot 2pq$$

$$\boxed{\text{Var} = \frac{pq}{2}}$$

Recall: $\Delta p = \frac{pq}{2\bar{w}} \frac{d\bar{w}}{dp}$ (Wright's equation)

Variance; Avg fitness; change in $\bar{w}$ with change in p

- Greater the variance, greater Δp i.e. more raw material
- Greater $\frac{d\bar{w}}{dp}$, greater Δp $\Rightarrow$ effect of selection
- Greater $\bar{w}$, smaller the Δp

Frequency-dependent selection

We'd assumed that w_{11}, w_{12} and w_{22} are independent of p. If we relax this, we get –

$$\frac{d\bar{w}}{dp} = 2pw_{11} + 2w_{12} - 4pw_{12} - 2w_{22} + 2pw_{22} + \dots$$

$$\frac{d\bar{w}}{dp} = 2\left(w_1^* - w_2^*\right) + E\left(\frac{dw}{dp}\right)$$

$\frac{dij}{dp}$ are multiplied by their frequencies

E: expectation

$$\Rightarrow \Delta p = \frac{pq}{2\bar{w}}\left(\frac{d\bar{w}}{dp} - E\left[\frac{dw}{dp}\right]\right)$$

$\frac{d\bar{w}}{dp}$ and $E\left[\frac{dw}{dp}\right]$ need not be equal / same – which gives rise to interesting dynamics.

"Selection always increases avg fitness" – need not be true

Relative fitness

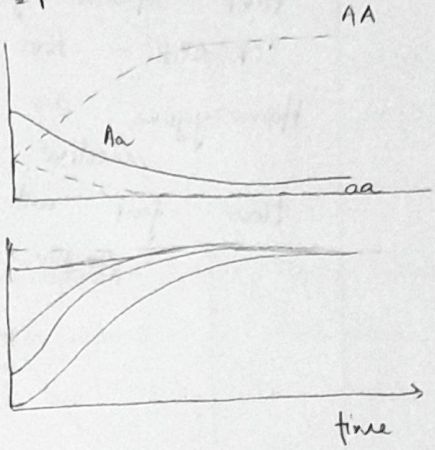
s: selection coefficient h: dominance coefficient

w_{AA}	w_{Aa}	w_{aa}	
1	1-hs	1-s	General dominance
1	1	1-s	selection against recessive
1	1-h	1	heterozygote disadvantage

If we use this instead of w_{11}, w_{12} & w_{22}, we'll get same result. h is positive

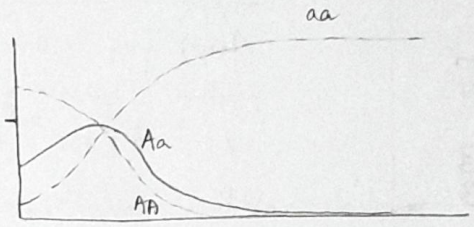
→ Selection against recessive: AA = Aa = 1 aa = 0.8

Allele freq change more rapidly in early generation when initial allele (A) is lower, because selectively favored dominant homozygote & heterozygote are relatively frequent in population



→ Selection against dominant phenotype: $w_{AA} = w_{Aa} = 0.8$ $w_{aa} = 1$

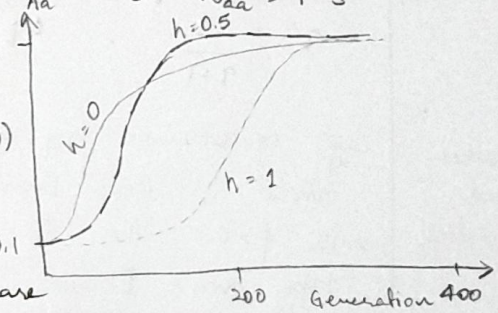
Heterozygote has lower relative fitness but its frequency initially increases because freq. of two alleles approach ~~equ~~ equality, but in the end, A goes extinct



Rate of change becomes faster as initial freq of A decreases

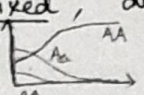
→ General case of dominance. $w_{AA} = 1$, $w_{Aa} = 1-hs$ $w_{aa} = 1-s$

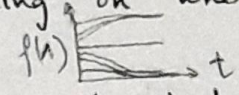
- Completely dominant A (h=0) - rapid change initially, but asymptotes later
- Completely recessive a (h=1) - slow initial change, then more rapid & slow fixation
- Additive effect of a (h=0.5) - intermediate case



→ Heterozygote disadvantage (Underdominance) $w_{AA} = w_{aa} = 1$ $w_{Aa} = 0.9$

Either allele gets fixed, depending on where we start

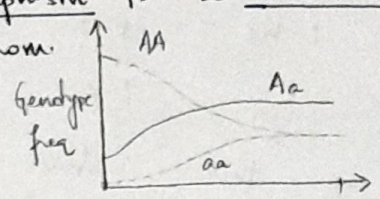


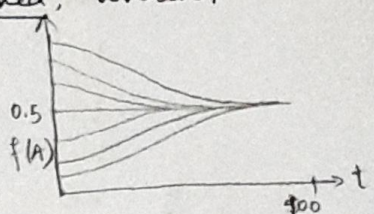


→ Heterozygote advantage $w_{AA} = 1 - s$ $w_{aa} = 1 - t$ $w_{Aa} = 1$

$$\text{Equilibrium} = \frac{t}{s+t}$$

We get stable polymorphism to be maintained, wherever you start from.





Empirical observations.

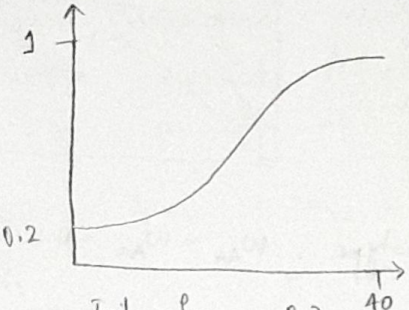
1. AIDS

HIV1 strain gains into the cell using CCR5 (cell surface protein)

CCR5Δ32 – loss of function mutation

Homozygons are resistant to HIV1 – should enjoy a selective pressure

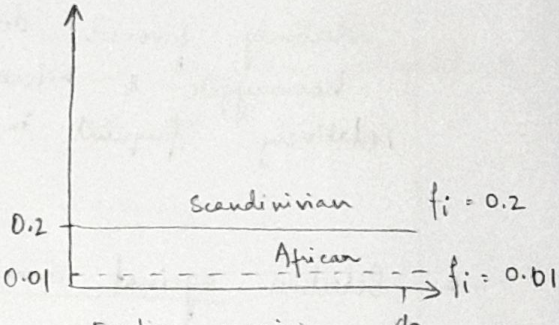
How fast will the freq. of CCR5Δ32 change?



Init. freq = 0.2

Fraction surviving:

+/+	+/Δ32	Δ32/Δ32
0.75	0.75	1



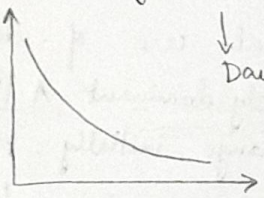
Fraction surviving – 40

(a)	0.995	0.995	1
(b)	0.75	0.75	1

Effect of selection: complex interplay of current allele frequency and selection pressure

2. Selection against recessive homozygous lethal (Tribolium)

$q' = \frac{q}{q+1}$; Δq reduces with time



Dawson et al

Overdominance leads to viable allele stabilizing at value < 1.

Mukai and Burdick

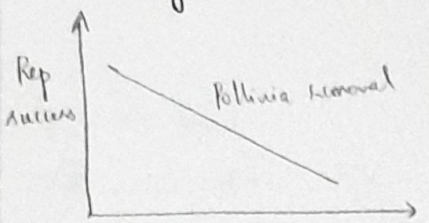
Say, equilibrium freq of viable allele is 0.8. Even if you start with $f > 0.8$, the freq equilibrates to 0.8. Why doesn't viable allele take over? Because fitness of heterozygote > dominant homozygote.

3. Color polymorphism in Elderflower orchids

Bumblebees alternate b/w yellow & purple ⇒ rarer flower colour should have better fitness

Negative frequency dependent selection –

They measured : pollinia being removed
pollinia being deposited & fruit set



Freq of yellow morph

12/4/22
~~11/4~~/22

Lecture

Mutation - Selection balance

- Involves taking a lot of approximations, but they work.
- Usually, the order of events is important in derivation, but the TB assumes it's not. Let's keep the order fixed –

Zygote (t) → Selection → Mutation → Zygote (t+1)

Ideally step 2 & 3 should be flipped

1 gene 2 locus: Assume A_1 is the allele under selection

Genotypic	A_1A_1	A_1A_2	A_2A_2
Fitness	$1-s$	$1-hs$	1
	$\frac{p^2(1-s)}{\bar{w}}$	$\frac{2pq(1-hs)}{\bar{w}}$	$\frac{q^2}{\bar{w}}$

h: degree of dom. coefficient

$h = 1 \Rightarrow A_1$ is dominant over A_2

$h = 0 \Rightarrow A_2 > A_1$

$0 < h < 1$ – intermediate fitness

$$\bar{w} = p^2(1-s) + 2pq(1-hs) + q^2$$

$$= p^2 - p^2 s + 2pq - 2hspq + q^2$$

$$= 1 - sp^2 - 2hs \cdot p(1-p)$$

A_1A_1	A_1A_2	A_2A_2
$\frac{p^2(1-s)}{\bar{w}}$	$\frac{2pq(1-hs)}{\bar{w}}$	$\frac{q^2}{\bar{w}}$

Freq of A_1 after selection – $p(p+q) = p$

$$p^s = \frac{p^2(1-s) + pq(1-hs)}{\bar{w}} = \frac{p^2 - sp^2 + pq - hspq}{\bar{w}}$$

$$p^s = \frac{p - sp^2 - hsp(1-q)}{\bar{w}}$$

Mutation : $A_1 \underset{\mu}{\overset{\gamma}{\rightleftharpoons}} A_2$

$$p' = p^s - \gamma p^s + \mu(1-p^s)$$

$$= p^s - \gamma p^s + \mu - \mu p^s$$

↳ ignore

$$\therefore p' \simeq p^s + \mu$$

Assumption (1)
Close to equilibrium, p^s is going to be very small
γ and μ are also v.v. small

$$\Rightarrow \Delta p = p' - p = p^s + \mu - p$$

$$= \frac{p - sp^2 - hsp(1-p) + \bar{w}(\mu - p)}{\bar{w}}$$

(2) Close to equilibrium
$\bar{w} \simeq w_{A_2A_2} \simeq 1$
because A_2A_2 dominates (as)

$$\Delta p = p - sp^2 - hsp(1-p) + \mu - p$$

$$\Delta p = \mu - sp^2 - hsp + hsp^2$$

★ When $h = 0$, $w_{A_2A_1} = w_{A_2A_2} = 1$ $\Rightarrow$ A_1A_1 is recessive lethal. This is a common special case.

$$\Rightarrow \Delta p = \mu - sp^2$$

At equilibrium : $\Delta p = 0$

$$\boxed{\hat{p} = \sqrt{\frac{\mu}{s}}}$$

Because we are taking √, the value of $\hat{p}$ is reasonably big

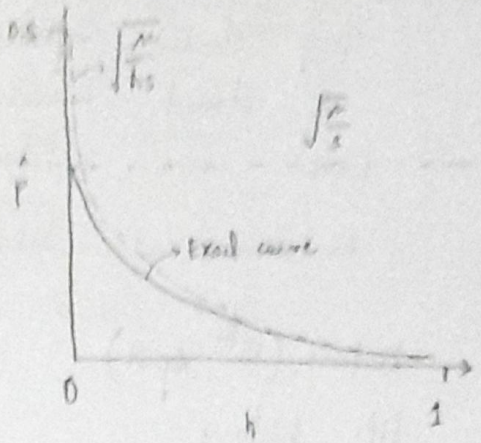
★ When $h \neq 0$. Ass. (3) : At equilibrium, $\hat{p}^2$ is v.v. small.

$$\Rightarrow \Delta p = \mu - hsp$$ ignoring p^2 terms.

At equilibrium : $\Delta p = 0$

$$\hat{p} = \frac{\mu}{hs}$$

The approximations give us values very close to exact values.



Genetic load

The departure b/w maximum avg. fitness and actual avg. fitness of population under specific conditions

* At $h=0$, $\bar{w} = 1 - sp^2$

load $= sp^2 = s \cdot \frac{\mu}{s} = \mu$

↳ $1 - \bar{w}$

⇒ What does this mean?

* $h \neq 0$ $\quad \bar{w} = 1 - sp^2 - 2hsp + 2hsp^2 \approx 1 - 2psh$

load $= 1 - \bar{w} = 2hsp = 2hs \cdot \frac{\mu}{hs}$

load $= 2\mu$

Load here is more than the previous case

Eg: Cystic fibrosis : Autosomal recessive condition - loss of mutation of CFTR gene which cause secretions to be very thick especially in the lungs.

This causes lung. sinus infections & infertility.

Here, $q = 0.02$

Assuming $s = 01$, we derive an estimation : $\mu = 4 \times 10^{-4}$

Realistic estimation : 6.7×10^{-7}. Why the huge difference in magnitude?

Lyczak 2002 found that Salmonella typhi might enter cell using CFTR. So, people with mutated CFTR might have lower vulnerability to typhoid.

They checked infectivity in mice - recessive mutant CFTR is almost immune to bacteria, ho and heterozygotes have some resistance than WT.

∴ Mutation and selection can interact in complex ways.

18/4

Lecture (13th April)

Life history

In nature, there's tremendous diversity in morphology

This diversity can be explained through natural selection

Eg. American plaice fish

Trait	Scotland	Newfoundland
Age at maturity	3 yrs	15 + yrs
Size at maturity	20 cm	40 cm
Longevity	6 yrs	20 + yrs
Maximum size	25 cms	50+ cms

How to explain this variation?

Life history is the set of all characters of an organism from its birth to death that have a bearing on its evolutionary fitness.

Eg: Size - at birth, at maturity, max
Sex ratios of offspring
Number of offspring
Age of maturity

Life history is all the axes along which phenotypes interact with environment to produce fitness.

The Darwinian Demon — with ideal life history —

- Spend min. time in development
- Start reproducing immediately after birth
- Produce infinite no. of offspring
- Live forever

Obviously, such an organism doesn't exist.

Constraints faced by the organism

1. Genetic constraints

Is there enough genetic raw material in the population?

eg. Sitona lineatus : major pest of legume plants

Native to Europe, where both winged & wingless forms are found. In New Zealand, only wingless form is found, although habitats are similar : founder effect — 'wing' allele wasn't available but it wasn't like that

What if trait is quantitative?

We need to look at heritability : high heritability ⇒ greater trait response

But, if a trait is under strong selection, then selection should erode additive genetic variance for that trait leading to low heritability. because of low variation

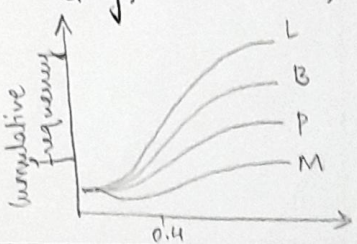
⇒ they don't respond to selection after sometime

Prediction : life history traits that will have the lowest heritability and shouldn't respond to selection

How to explain low variation in quantitative traits?

Mousseau & Roff 1987

- 1120 estimates from natural populations
- Divided into 4 categories — life history, behaviour, physiology & morphology
- Expectation : LH traits would have lowest heritability



The graph tells us that maximum variation is found in morphological traits, and least in LH traits. (↓ herit)

A large fraction of LH traits actually has high heritability.

⇒ the heritability is not so low that selection can act on it.

It's possible to have enough standing variation in natural populations for LH-traits to elicit rapid response from to selection.

Till the '80s, it was believed that selection can't act on conserved traits anymore.

- Eg. Drosophila development

Drosophila mature on rotting fruit, which is unpredictable and ephemeral. So there's an enormous pressure on larvae to develop ASAP.

It was thought that drosophila development time has been minimised & couldn't be reduced.

But, experiments have shown that development time can be reduced still further.

But, with fast development, they have small size, short longevity, low fecundity and have deformations.

- What maintains genetic variation?

1. Mutation-selection balance
2. Heterozygote advantage
3. Frequency dependent selection – keeps dynamic
4. In a heterogenous environment.

2. Phylogenetic constraints

Since evolution is a historical process, there are phylogenetic constraints. In a sense, it's an extension of genetic cons

Eg: Emperor penguins & kiwis — are wingless. Even if they're strongly selected for flight, the chances of regaining wings are very low.

In some cases, phylogenetic constraints can be overcome.

Eg: Flat fish — they live at the bottom of sea, so they swim flatly, so their eyes lie on same side & so on. A species from the same clade doesn't swim at the bottom — its eye has migrated back to the top of its head, which is a huge change

3. Physiological Constraints.

Constraints that result from proceses internal to the organism, but above the level of the gene

Eg: • Smaller the body size, higher the metabolic rate
Proposed lower limit of mammal size is 2.25 g

• Offspring size of organisms wrt size of reproductive tract. Human head size - birth canal

4 Mechanical constraints

Eg: Gravity regulates the height of giraffe's neck and largest mammals are found in water (buoyancy)

5. Ecological constraints

Eg predators, pathogens etc.

Given that many constraints influence phenotype simultaneously, how does one figure out what is responsible for what?

Often, many bito biologists make the <u>adaptationist argument</u> – the trait must contribute to fitness – without actually checking.

Lewontin & Gould : Spandrels of San Marco.

Adaptationist Program – "it's there for a reason"

Eg: <u>Oxpecker</u> is a bird that is found in association with large mammals, feeding off the insects, but sometimes even pecks the wound, and might not significantly reduce pathogen load.

If something is potentially adaptive, it doesn't mean that it has to be

19/4/22

Lecture

Offspring number

Clutch size : no. of ~~rep~~ offspring in a single reproductive effort.

<u>Semelparous</u> – organisms that produce a large no. of offspring in one bout and die

<u>Iteroparous</u> – organisms that reproduce several offspring in reproductive bouts over its lifespan

Factors –

→ Stability of environment : r vs k selection logic assumes <u>survivorship - fecundity tradeoff</u>. For organisms in unpredictable environment should produce <u>constant offspring</u> – <u>geometric mean matters</u>

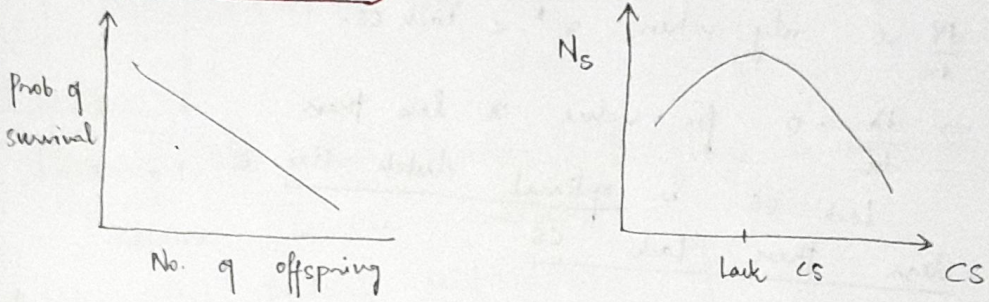
→ Offspring and adult size - physiological constraints

→ Resource availability and intraspecific competition - which determines offspring survivorship.

Organisms should produce no. of offspring, so per capita resource is sufficient, and survivorship is maximised - Lack Clutch size.

David Lack - selection favors clutch size that maximizes no. of surviving offspring.

$N_s = CS \times P_{cs}$ $\quad P_{cs} = 1 - m \times CS$



Empirical test of Lack's hypothesis - Boyce & Perrins 1987

Study on Great tits

The mean clutch size was 8-9, but maximum survival was for CS = 12.

They manipulated clutch size, but maximum survival was still for CS = 12. So, they can technically care for more young successfully.

Many other studies showed similar patterns. Why would this be happening?

In calculating Lack clutch size, we ignore the cost to the mother/parents.

(64)

Consider –

$$N_{t+1} = SN_t + \alpha x N_t$$

functions of x { S: adult survivorship
α: offspring survivorship
x: no. of offspring

$$N_{t+1} = N_t (S + \alpha x)$$

$$N_{t+1} = \lambda N_t \quad \text{where} \quad \lambda = S + \alpha x = S + Y$$

To find maxima –

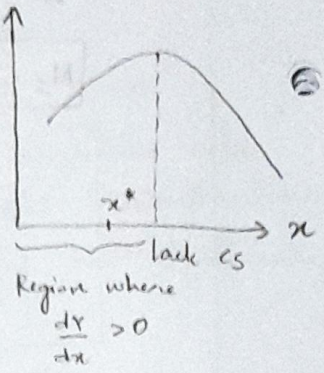
$$\frac{d\lambda}{dx} = \frac{dS}{dx} + \frac{dY}{dx} = 0 \quad \Rightarrow \quad \frac{dS}{dx} = -\frac{dY}{dx}$$

If we assume S decreases with clutch size, the value of $\frac{dS}{dx}$ will be negative.

$\Rightarrow \frac{dY}{dx}$ should be positive.

$\frac{dY}{dx} > 0$ only when $x^* <$ Lack cs. $Y = \alpha x$

$\Rightarrow \frac{d\lambda}{dx} = 0$ for value x less than Lack cs. So optimal clutch size is less than Lack cs

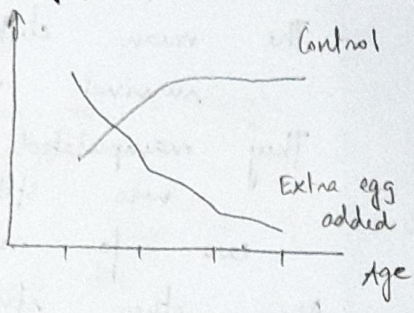


1. Does reproductive effort in later life tradeoff with present clutch?

Gustaffson & Part 1990 – collard flycatchers

They added one extra egg to the nest at age 1

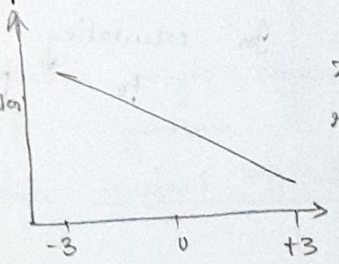
Females forced to raise extra offspring in early life pay a cost later in life



2. What is there are other costs to the offspring?

Schluter & Gustaffson 1993 – they followed clutch size of daughters whose mother's clutch size were manipulated

⇒ There's a quantity-quality tradeoff.



Lack's hypothesis : More discussion

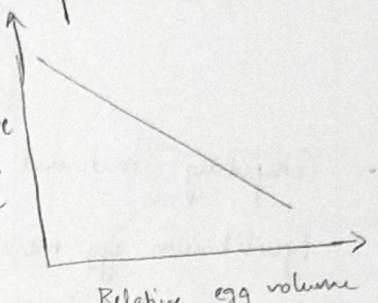
- If clutch size varies with genotype and there's variation, then modal CS will be different than optimal CS
- If there's environmental heterogeneity, then optimal CS will be less than Lack CS
- ∴ Lack CS is <u>useful as a null model</u>, rather than an accurate descriptor.

What if each offspring is not the same?
Offspring size and number trade-off.

Lecture 20/4

Offspring size and number : few large or numerous small?
Because resource available is finite

Does this trade-off exist? Yes!
It has been observed in 26 fish families and Drosophila

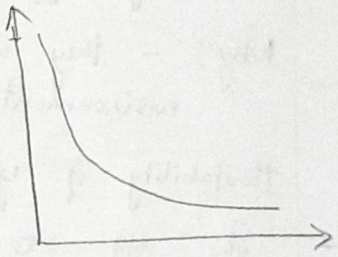


What is an optimal strategy?

Smith & Fretwell 1974 Model

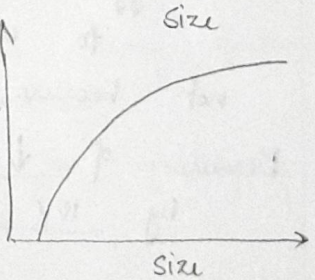
Assumption 1: <u>Tradeoff b/w offspring size and number</u>

Here, $10/size$



Assumption 2: Above a threshold, <u>offspring survival increases and then saturates w/ size</u>

Here, $1 - \frac{1}{size}$



(56)

Combining the two vogel —

So parental fitness is maximised for an intermediate size

But for offspring, maximal fitness = maximum size

This leads to parent offspring conflict

Empirical study: Heath et al 2013.

Chinook salmon or king salmon — very big fish that is prized for sport & food and endangered

Salmons hatch in stream, travel to sea and return to the same stream to lay its eggs

There are artificial commercial hatcheries which do external insemination, which helps in ex-situ conservation.

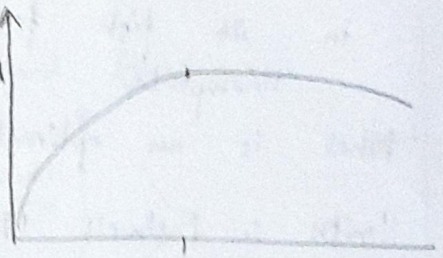
Testing assumptions: There's a tradeoff ie. bigger clutch size means smaller eggs.

* Offspring survival increases with size, not a strong relationship though.

Computing maternal fitness

Equilibrium egg mass: 0.16g

But in the beginning of study, egg mass = 0.27g



Why? — they are in a benign environment ⇒ no mortality, resource competition & environmental fluctuations.

Heritability of egg mass = 0.26

So, egg size in hatchery should reduce, and go closer to 0.16g. This is what happened and its not because of female body size.

Because of ↓ in egg mass, mean fecundity went up by 10% in 4 years, which is very high.

If heritability > 0.2, strong effect of selection

Impact on conservation

These people also seeded fish in rivers after counting/estimating natural population size and carrying capacity.

The supplemented population (with lower egg size) didn't fare so well – maternal fitness (fecundity x survival) and egg survival decreased.

So, captive breeding changes the genetic make-up of the population, which can have deleterious effects on life-history and hence the quality of re-introduced individual.

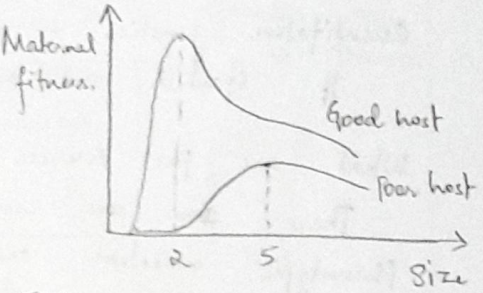
Do organisms have a choice? Fox et al 1997

They looked at A the seed beetle ~~feed~~ lays eggs on many hosts, two of which are –

Catclaw acacia : good host, high survivorship

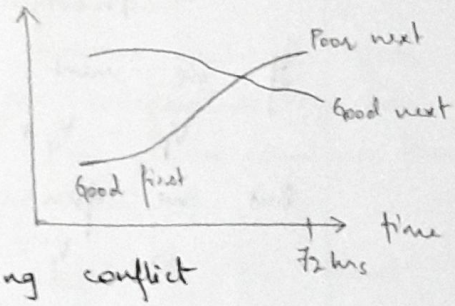
Blue palo verde : harsh host, low survivorship.

As expected, organisms lay larger eggs on poor host and vice versa.



Host-switch experiment

Allow females to lay eggs on one host and then transfer to different host. There is plasticity



Take-home messages

- Selection for offspring size vs no. can be shaped by parent-offspring conflict
- Exact relation ~~ship~~ will depend on many factors, including offspring size & survival
- Life history traits are dependent on quality of environment

If life-history evolution always for the better?

FET-JB

Going back to fast-developing Drosophila — their development is very fast, but they are deformed, have low energy, die faster.

So, selection is very potent, but also very specific. Which means, when you're selecting for one thing, you might also be selecting for something else inadvertently. For instance, when they were selecting for longer lifespan, they also selected for late life fecundity.

So, in breeding programs, you have to be very careful about the direction in which we're leading the population.

Lecture 06/4

Introduction to Quantitative genetics → simplest way

3 strands to evolutionary modelling — population genetics, quantitative genetics & evolutionary game theory

Quantitative genetics — oldest and most powerful way. It started out with RA Fisher's 1918 paper.

What are the sources of variation in traits?

There are continuous traits — height, weight etc.

Phenotypic variation can come from 2 sources —

Phenotypic variation = Genetic var. + Environmental var

If we want to write this in terms of variances —

$$V_P = V_G + V_E + 2\,Cov(G,E)$$

But we often assume that there's no G×E interaction ⇒ Cov = 0

$$\Rightarrow V_P = V_G + V_E$$

Contribution to phenotypic variance that derives from genotypic variance is transmitted genetically from parent to offspring

Contribution from V_E is not transmitted.

Broad sense heritability: (fraction of variance from gene)

$$H^2 = \frac{V_G}{V_G + V_E}$$

not a super-useful quantity.

Measuring H^2

1. Using highly inbred lines
   - Assume they have so little genetic variation that they can be treated as genotypically homogenous
   - ⇒ phenotypic variance within individuals of a line is purely environmental (estimate of V_E), while phenotypic variance across individuals from various lines is an estimate of V_P ($V_G + V_E$)

   ⇒ $V_G = V_P - V_E$

   But this is difficult to carry out for all organisms

2. Compare monozygotic & dizygotic twins

   Monozygotic – genetically identical ⇒ only V_E

   Dizygotic – like siblings ⇒ $V_E + V_G$

   Eg: (susceptibility)$_x$ ∝ occurrence of depression is greater in MZ twins than DZ twin pairs. ⇒ there's a genetic susceptibility to depression.

Further decomposition of V_G

$V_P = V_G + V_E$

$$V_G = V_A + V_D + V_I$$

V_A : additive component → effect of allele, independent of background

V_D : dominance component → other alleles at same locus

V_I : interactive / epistatic component → effect due to alleles at other loci

assumes that all effects are independent of each other ⇒ no covariance terms

Say A & B alleles increase the height of the plant

A – +0.1m ; B : +0.3m

	aabb	Aabb	AaBb	AABB	
Additive	1	1.1	1.4	1.8	
Dominance	1	1.2	1.8	1.8	A or AA : +0.2m ; B or BB : +0.6m ⇒ complete dominance
Interactive	1	1.1	1.8	1.8	A : +0.1 ; B : +0.3 m ; A/B : +0.2m ; B/A : +0.6 m

(10)

Narrow - sense heritability : h^2

A, D & I contribute to V_G. D & I are dependent on genetic background. D is not heritable; and inheritance of I is dicey.

V_A is context dependent ⇒ more accessible to nat. sel

So we look at the fraction of total variation that is due to additive genetic variation alone —

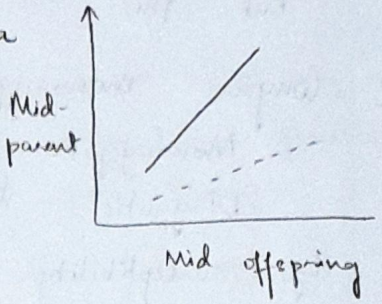
$$h^2 = \frac{V_A}{V_A + V_D + V_I + V_E}$$

When 'heritability' is used, people mean h^2.

Measuring h^2

h^2 is the slope of regression of offspring phenotype against mid-parent value. Midparent = Avg of mom + dad

Eg. Heritability of beak depth in geospiza



Issues in estimating heritability —

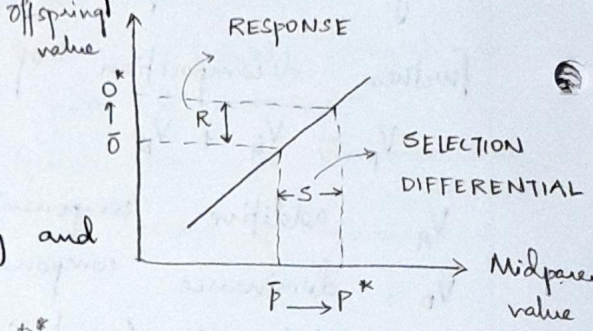
- Misidentified paternity
- Shared environments
- Maternal effects

~~Some notes on heritability~~

~~Heritability~~ Breeder's Equation : Response to selection

$$\text{Slope} = h^2 = \frac{O^* - \bar{O}}{P^* - \bar{P}} = \frac{R}{S}$$

$$\Rightarrow \boxed{R = h^2 S}$$



∴ Multiplicative effect of heritability and selection strength

Eg. Abdominal bristle no. in fruit flies P^* - mean value of subset of individuals selected

1. Estimate h^2 value = 0.51 (using parent offspring curve)
2. In some population, select higher bristle no. parents and breed to make next generation

   $S = 40.6 - 35.3 = 5.3$

3. $\Theta^* = 37.9$, $\bar{\Theta} = 35.3$ (value remains same as prev. generation)

   $\Rightarrow R = 2.6 \Rightarrow h^2 = 0.49$ (close to the regression value).

Lecture

Some notes on heritability (h^2)

Heritability measures the portion of parental population phenotypic variance that is passed on to the offspring generation through gametes

- Heritability is a population measure, and it's explicitly a function of allele and genotypic frequencies
- If any population level attributes are altered, heritability can change
  ⇒ h^2 is defined for a given population in a particular context
- A population would only respond to selection if there exists selection pressure and heritability of trait.

If there's no additive genetic variation, $h^2 = 0$ ⇒ no matter the value of S, $R = 0$.

This is additivity at population level; dominance & epistasis at Mendelian level can also contribute to selection

- h^2 has V_E in denominator. Any statement about h^2 has to be qualified by the environment in which it is measured
- Critical ex: heritability of IQ across races (IQ is a terrible metric)
  But IQ has a high h^2, especially among white people
  Challenge: environment as a matter of socio-economic status
  White people are better off.
- Another eg: Plant height – 7 genotypes planted in 2 cities
  h^2 is very high. Strongly differing heights b/w → Same genotype, high h^2, different places leads to different heights solely due to different environments
  Env. plays important role, especially for plastic traits. ∃ a genetic basis, but resource quality ↑ growth