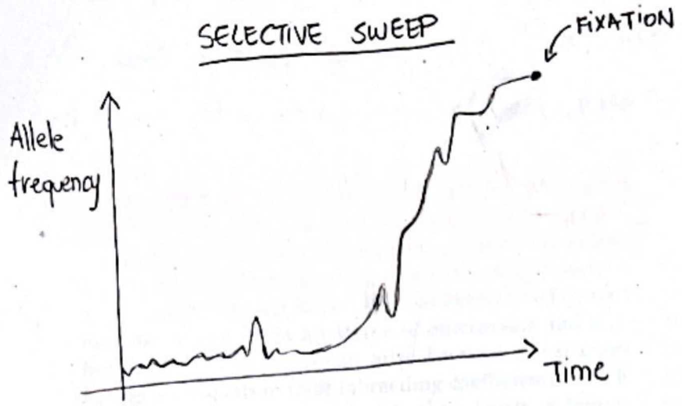
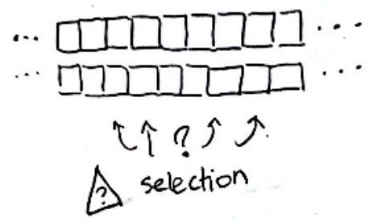
SELECTIVE SWEEP



FIXATION
- Allele is positively selected: $E[\Delta p | p] > 0$.
- Stochasticity: $Var[\Delta p | p] >> 0$.

Look at Genomic data



Metrics ⟹ VARIABILITY

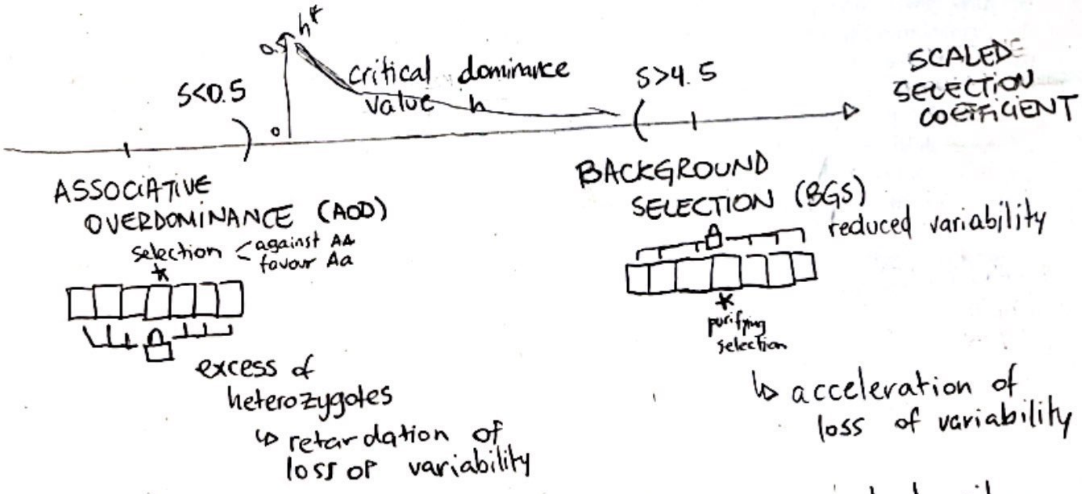
π - nucleotide diversity

θ_ω - Watterson theta

} D - Tajima D

$\frac{dN}{dS}$, $\frac{K_a}{K_s}$ ratio, $\omega(\delta)$ — nonsynonymous / synonymous Ortholog substitutions

⚠ Metrics are sensitive to linkage between loci.

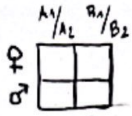


GOAL: understand the conflict of AOD and BGS on neutral sites.

POPULATION GENETICS MODEL — two diallelic loci in diploids [Ewens, pg. 67]

well-mixed

WF life-cycle



*

FITNESS
$$\begin{cases} \omega(B_1B_1) = 1 \\ \omega(B_1B_2) = 1 - hs \\ \omega(B_2B_2) = 1 - s \end{cases}$$
Purifying selection ($s>0$, B_2 deleterious)

MUTATION
$$\begin{cases} B_1 \to B_2 : u \\ B_2 \to B_1 : v \end{cases}$$

HAPLOTYPE FREQUENCIES
$$\begin{cases} A_1B_1 : y_1 \\ A_1B_2 : y_2 \\ A_2B_1 : y_3 \\ A_2B_2 : y_4 \end{cases}$$
$y_1 \oplus y_2 \to x : A_1$; $y_1 \oplus y_3 \to q : B_2$ — ALLELE FREQUENCIES

RECOMBINATION : c $\quad (0 < c \leq 0.5)$

$\uparrow$ superlocus $\qquad \uparrow$ different chromosomes

a) If both loci are neutral ($s=0$):

HAPLOTYPE FREQ. CHANGE
$$\begin{cases} y_1' = y_1 - c\,[y_1y_4 - y_2y_3] \\ y_2' = y_2 + c\,[y_1y_4 - y_2y_3] \\ y_3' = y_3 + c\,[y_1y_4 - y_2y_3] \\ y_4' = y_4 - c\,\underbrace{[y_1y_4 - y_2y_3]}_{\text{coefficient of linkage disequilibrium } (D)} \end{cases}$$

$(D) \uparrow$ eigenfunction

$$A_1B_1 \xrightarrow{c} A_1B_2$$
$$A_2B_2 \xleftarrow{c} A_2B_1$$

$$D' = y_1'y_4' - y_2'y_3' = (1-c)[y_1y_4 - y_2y_3] = (1-c)D$$

⚠ D exponentially decays to 0 if $c \neq 0$.

b) If B is under selection

♂ \ ♀	A_1B_1	A_1B_2	A_2B_1	A_2B_2
A_1B_1	1	1-hs	1	1-hs
A_1B_2	1-hs	1-s	1-hs	1-s
A_2B_1	1	1-hs	1	1-hs
A_2B_2	1-hs	1-s	1-hs	1-s

→ MARGINAL FITNESS OF GAMETE
$$\omega_i = \sum_{j=1}^{4} y_j \omega_{ij}$$

→ MEAN FITNESS OF POPULATION
$$\bar{\omega} = \sum_{i,j=1}^{4} y_i y_j \omega_{ij}$$

HAPLOTYPE FREQ CHANGE

$$\begin{cases} y_1' = \dfrac{y_1 w_1 - c(1-hs)\left[y_2 y_3 - y_1 y_4\right]}{\overline{w}} \\ y_2' = \overline{w}^{-1}\left\{y_2 w_2 + c(1-hs)\left[y_2 y_3 - y_1 y_4\right]\right\} \\ y_3' = \overline{w}^{-1}\left\{y_3 w_3 + c(1-hs)\left[y_2 y_3 - y_1 y_4\right]\right\} \\ y_4' = \overline{w}^{-1}\left\{y_4 w_4 - c(1-hs)\left[y_2 y_3 - y_1 y_4\right]\right\} \end{cases}$$

⚠ Recombination happens in heterozygotes $\left[c(1-hs)\right]$. (Link Overdominance & Recombination)

💡 IDEA: Although locus A is itself unselected, it'll appear to have a selective effect due to its association with B locus.

→ APPARENT FITNESS OF NEUTRAL LOCUS:

$$\left.\begin{aligned} \overline{w}_{A_1A_1} &= \frac{y_1^2 + 2y_1y_2(1-hs) + y_2^2(1-s)}{y_1^2 + 2y_1y_2 + y_2^2} \\ \overline{w}_{A_1A_2} &= \frac{y_1y_3 + (y_2y_3 + y_1y_4)(1-hs) + y_2y_4(1-s)}{y_1y_3 + y_2y_3 + y_1y_4 + y_2y_4} \\ \overline{w}_{A_2A_2} &= \frac{y_3^2 + 2y_3y_4(1-hs) + y_4^2(1-s)}{y_3^2 + 2y_3y_4 + y_4^2} \end{aligned}\right\} \text{Eqs. 1a – 1c.}$$

→ APPARENT SELECTION COEFFICIENTS AGAINST HOMOZYGOTES

$$\tilde{s} = \mathbb{E}\left[\tilde{W}_{A_1A_2} - \tilde{W}_{A_1A_1}\right] = \frac{-(1-x)\left[x^2 - 2y_1y_2hs - y_2^2 s\right] + x\left[x(1-x) - y_1y_4hs - y_2y_3hs - y_2y_4s\right]}{x^2(1-x)}$$

$$= s\,\mathbb{E}\left[-\frac{D\left[h + (y_2+y_4)(1-2h)\right]}{(y_1+y_2)(y_3+y_4)} + \frac{D^2(1-2h)}{(y_1+y_2)^2(y_3+y_4)}\right]$$

$$\tilde{t} = \mathbb{E}\left[\tilde{W}_{A_1A_2} - \tilde{W}_{A_2A_2}\right] = s\,\mathbb{E}\left[+\frac{D\left[h + (y_2+y_4)(1-2h)\right]}{(y_1+y_2)(y_3+y_4)} + \frac{D^2(1-2h)}{(y_1+y_2)^2(y_3+y_4)}\right]$$

⚠ If at $t=0$, $\mathbb{E}[D]=0$ (no linkage disequilibrium), then $\mathbb{E}[D(t)]=0$, and $\mathbb{E}[D(y_2+y_4)^n]=0$.

$$\tilde{s} \simeq s(1-2h)\,\mathbb{E}\left[\frac{q(1-q)}{x} \times \frac{D^2}{x(1-x)q(1-q)}\right] \qquad \left(q = \text{freq of } B_2\right)$$

$$\tilde{t} \simeq s(1-2h)\,\mathbb{E}\left[\frac{q(1-q)}{1-x} \times \frac{D^2}{x(1-x)q(1-q)}\right]$$

and

$$R^2 = \frac{D^2}{x(1-x)q(1-q)}$$ squared correlation coefficient in allelic state between A_1 and B_2

$\Rightarrow$ Both $\tilde{s}$ & $\tilde{t}$ are positive: selection favours heterozygotes.

To calculate $\tilde{s}$ & $\tilde{t}$: assume independence of x, q, r.

a) Inbred load, $B = s(1-2h)q(1-q)$

$$B = \log(W_{F=0}) - \log(W_{F=1})$$

$$= \log\left((1-q)^2 + 2q(1-q)(1-hs) + q^2(1-s)\right) - \log\left((1-q) + q(1-s)\right)$$

$$= \log(1 - 2hsq(1-q) - sq^2) - \log(1-qs)$$

$\boxed{\log(1+\varepsilon) = \varepsilon}$

$$= -2hs\,q(1-q) - sq^2 + sq = s(1-2h)q(1-q)$$

Equilibrium: $B^* = (1-q^*)q^* \cdot s(1-2h)$ with $q^* = \dfrac{u - v}{hs}$ neglect.

$$B^* = \left(1 - \frac{u}{hs}\right)\frac{u}{h\cancel{s}} \cdot \cancel{s}(1-2h)$$

$$= \left(1 - \frac{u}{hs}\right)\left(\frac{2u}{2h} - 2u\right) \simeq 2u\left(\frac{1}{2h} - 1\right)$$

b) Approximate Neutral Recursion for $E\{r^2\}$

$D = y_2 y_3 - y_1 y_4$

A_1B_1: $y_1 = p_{A_1} p_{B_1} + D \rightarrow y_1 - \bar{y}_1 = D$

A_1B_2: $y_2 = p_{A_1} p_{B_2} - D \rightarrow y_2 - \bar{y}_2 = -D$

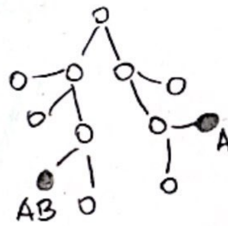
A_2B_1: $y_3 = p_{A_2} p_{B_1} - D \rightarrow y_3 - \bar{y}_3 = -D$

A_2B_2: $y_4 = p_{A_2} p_{B_2} + D \rightarrow y_4 - \bar{y}_4 = D$

$$r^2 = \frac{D^2}{p_{A_1} p_{A_2} p_{B_1} p_{B_2}} = \frac{cov(A,B)}{\sqrt{Var(A)} \times \sqrt{Var(B)}}$$ PEARSON CORRELATION COEFFICIENT

PROBABILITY OF COALESCENCE

Take two A-locus genes, that are identical by descent.
Then, the genes at the B-locus could be IBD simply through there having been no recombination between the two loci on either of the pathways from the common anscestor.



Let Q be the probability of no recombination.

Q = conditional probability of joint IBD.

SVED 1971: $$\boxed{Q = \frac{D^2}{p_{A_1} p_{A_2} p_{B_1} p_{B_2}} = r^2}$$

RECURSION FOR Q:

$$Q_{t+1} = \underbrace{\frac{1}{2N}}_{\text{from the same gene in the previous generation}} \underbrace{(1-c)^2}_{\text{no recombination}} + \underbrace{\left(1 - \frac{1}{2N}\right)}_{\neq \text{ genes that coalesce}} Q_t \underbrace{(1-c)^2}_{\text{no recombination}}$$

If $c << 0.5$ & neglect c/N:

$$Q_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N} - 2c\right) Q_t \qquad (A4a)$$

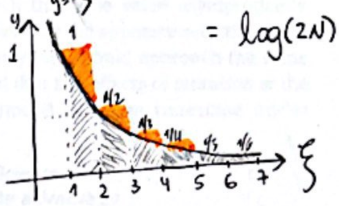
$$Q^* = \frac{1}{1-4Nc} \; \underset{\text{including mutations}}{=} \; \frac{1}{1-4N(c+u+v)}$$

c) Expectations of frequency distribution $\mathbb{E}\{x^{-1}\}, \mathbb{E}\{(1-x)^{-1}\}$

The conditional probability distribution becomes uniform with time, under pure drift regardless initial frequency.

$$\mathbb{E}\{x^{-1} \mid t \gg 2N\} = \sum_{i=1}^{2N-1} \frac{1}{i/2N}\cdot\frac{1}{2N} = \sum_{i=1}^{2N-1}\frac{1}{i} \simeq \int_1^{2N}\frac{1}{\xi}\,d\xi + \text{EXTRA BITS}$$

$$= \log(2N) + \gamma$$



$\uparrow$ Euler-Mascheroni constant

Because the sum is symmetric $i \leftrightarrow 2N-i$, $\mathbb{E}\{(1-x)^{-1} \mid t \gg 2N\} = \log(2N)+\gamma$

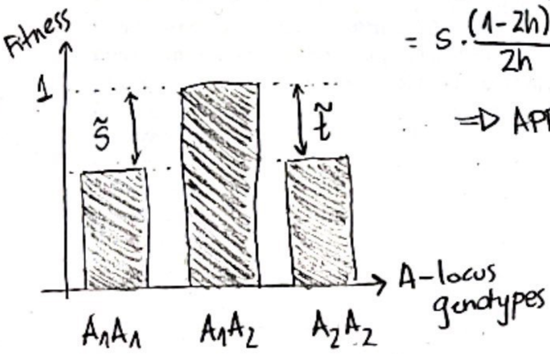
—//—

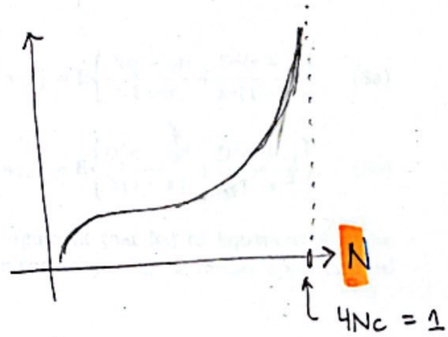
Putting a), b), c) together

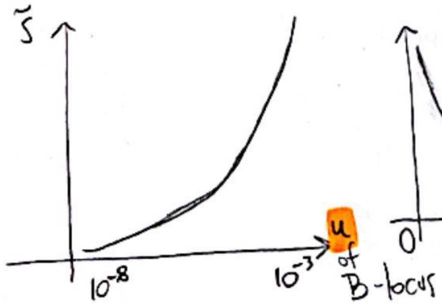
$$\tilde{s} = \tilde{t} = s(1-2h)\left[2u\left(\frac{1}{2h}-1\right)\right]\cdot\left[\frac{1}{1-4N(c+u+v)}\right]\left(\log(2N)+\gamma\right)$$

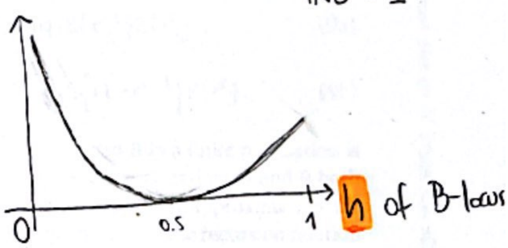
$$= s\cdot\frac{(1-2h)^2}{2h}\cdot\frac{2u(\log(2N)+\gamma)}{1-4N(c+u+v)}$$

$\Rightarrow$ APPERENT HETEROZYGOTE AT A-LOCUS.









—//—

If Fitness $\begin{cases} w(B_1B_1): 1-s \\ w(B_1B_2): 1 \\ w(B_2B_2): 1-t \end{cases}$ (overdominance)

$$\tilde{s} \simeq (s+t)\,\mathbb{E}\{pq\}\,\mathbb{E}\{x^{-1}\}\,\mathbb{E}\{r^2\}$$

$$\tilde{t} \simeq (s+t)\underbrace{\mathbb{E}\{pq\}\,\mathbb{E}\{(1-x)^{-1}\}\,\mathbb{E}\{r^2\}}_{\text{EQUAL AS BEFORE}}$$

c) What are we keeping track of?

Heterozygosity at neutral site: $H = 2\,\mathbb{E}\{x(1-x)\}$

$$\Delta H = \frac{dH}{dx}\Delta x \implies \Delta H = 2(1-2x)\Delta x + O(\Delta x^2)$$

$$\mathbb{E}\{\Delta H\} = 2s\,\mathbb{E}\{D(1-2x)[h + q(1-2h)]\}.$$

//

How to calculate the rate of loss of variability?

Linear diffusion operator

$$f(x,q,D) = f(x^*,q^*,D^*) + \partial_x f(x^*,q^*,D^*)(x-x^*) + \partial_q f(x^*,q^*,D^*)(q-q^*) + \partial_D f(x^*,q^*,D^*)(D-D^*) + \tfrac{1}{2}\partial_x^2 f(x^*,q^*,D^*)(x-x^*)^2 + \partial_{xq}^2 f(x^*,q^*,D^*)(x-x^*)(q-q^*)$$

$$\Delta\mathbb{E}\{f\} = \overbrace{\mathbb{E}(x-x^*)}^{M_x = \mathbb{E}[x]}\partial_x f^* + \overbrace{\mathbb{E}(q-q^*)}^{M_q}\partial_q f^* + \mathbb{E}(D-D^*)\partial_D f^* + \underbrace{\tfrac{1}{2}\mathbb{E}\{(x-x^*)^2\}}_{V_x = \mathbb{E}[x^2]-\mathbb{E}[x]^2}\partial_x^2 f^* + \underbrace{\mathbb{E}\{(x-x^*)(q-q^*)\}}_{C_{xq}}\partial_{xq}^2 f^*$$

Ohta & Kimura, 1971:

$$\begin{bmatrix} x(1-x) \\ x(1-x)q \\ x(1-x)q^2 \\ (1-2x)D \\ (1-2x)Dq \\ (1-2x)Dq^2 \\ D^2 \\ D^2q \\ D^2q^2 \end{bmatrix}_{t+1} = R_{9\times 9} \begin{bmatrix} x(1-x) \\ x(1-x)q \\ x(1-x)q^2 \\ (1-2x)D \\ (1-2x)Dq \\ (1-2x)Dq^2 \\ D^2 \\ D^2q \\ D^2q^2 \end{bmatrix}_{t}$$

OBS:

$$\mathbb{E}\{r^2\} = \frac{Y_7}{Y_2 - Y_3}$$

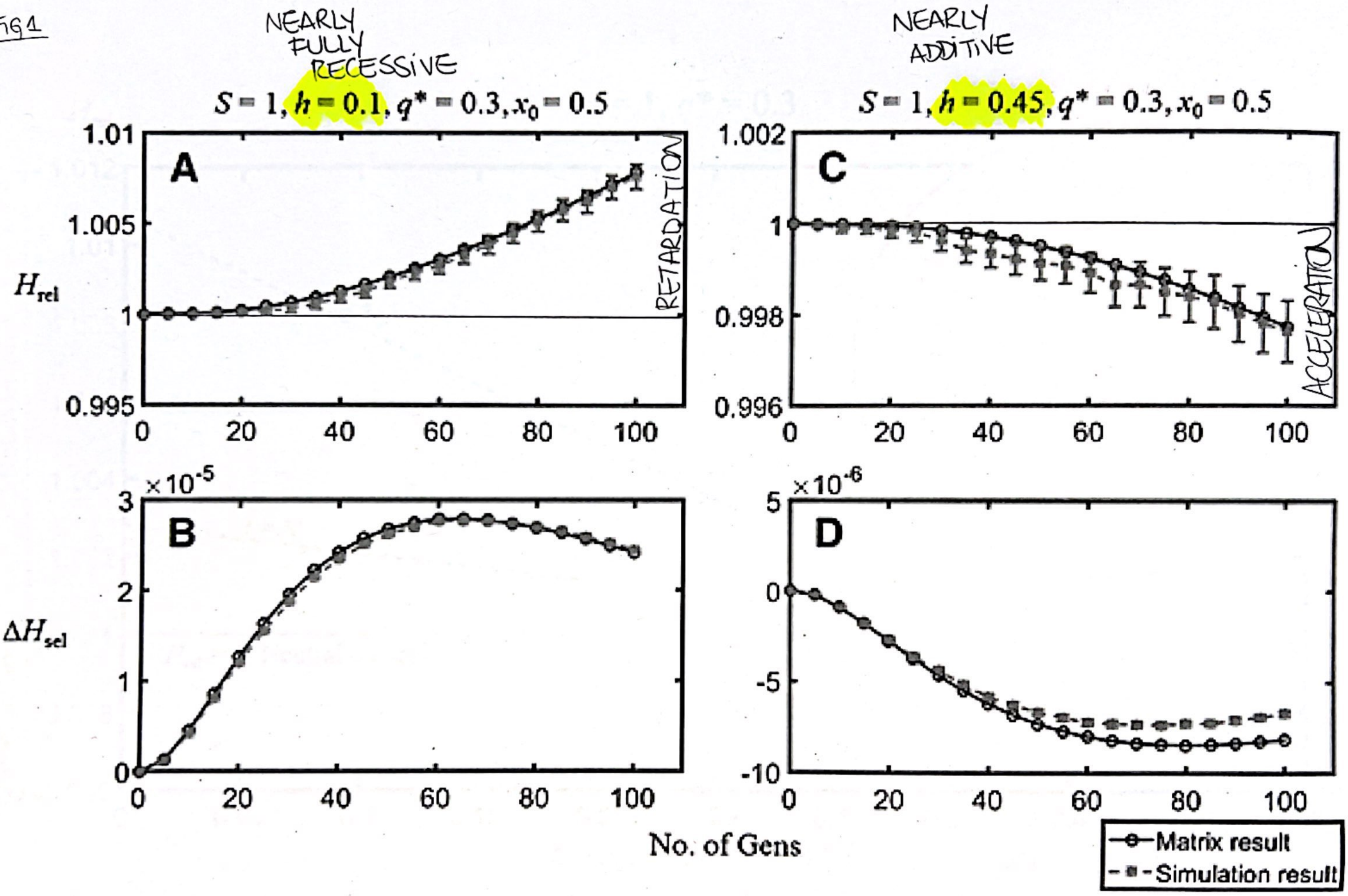
$$\mathbb{E}[H] = 2Y_1 = H_\infty = -2\lambda_0^{-1}\left(\frac{2uv}{M+v}\right)$$

//

Relative heterozygosity: $H_{rel} = \dfrac{x(t)[1-x(t)]}{x_0(1-x_0)}$ → no selection

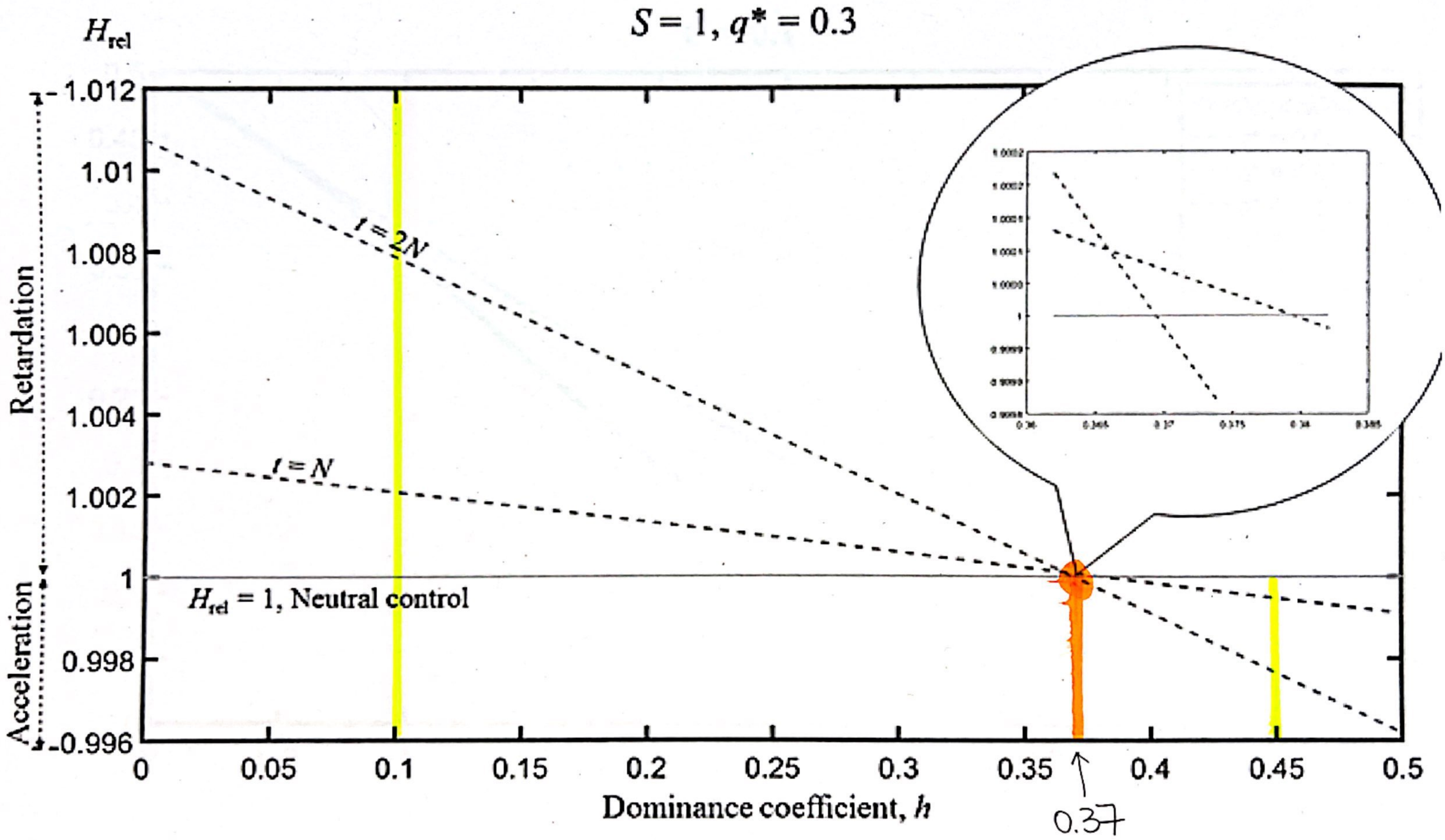
& $D_0 = 0$.

Fig 1



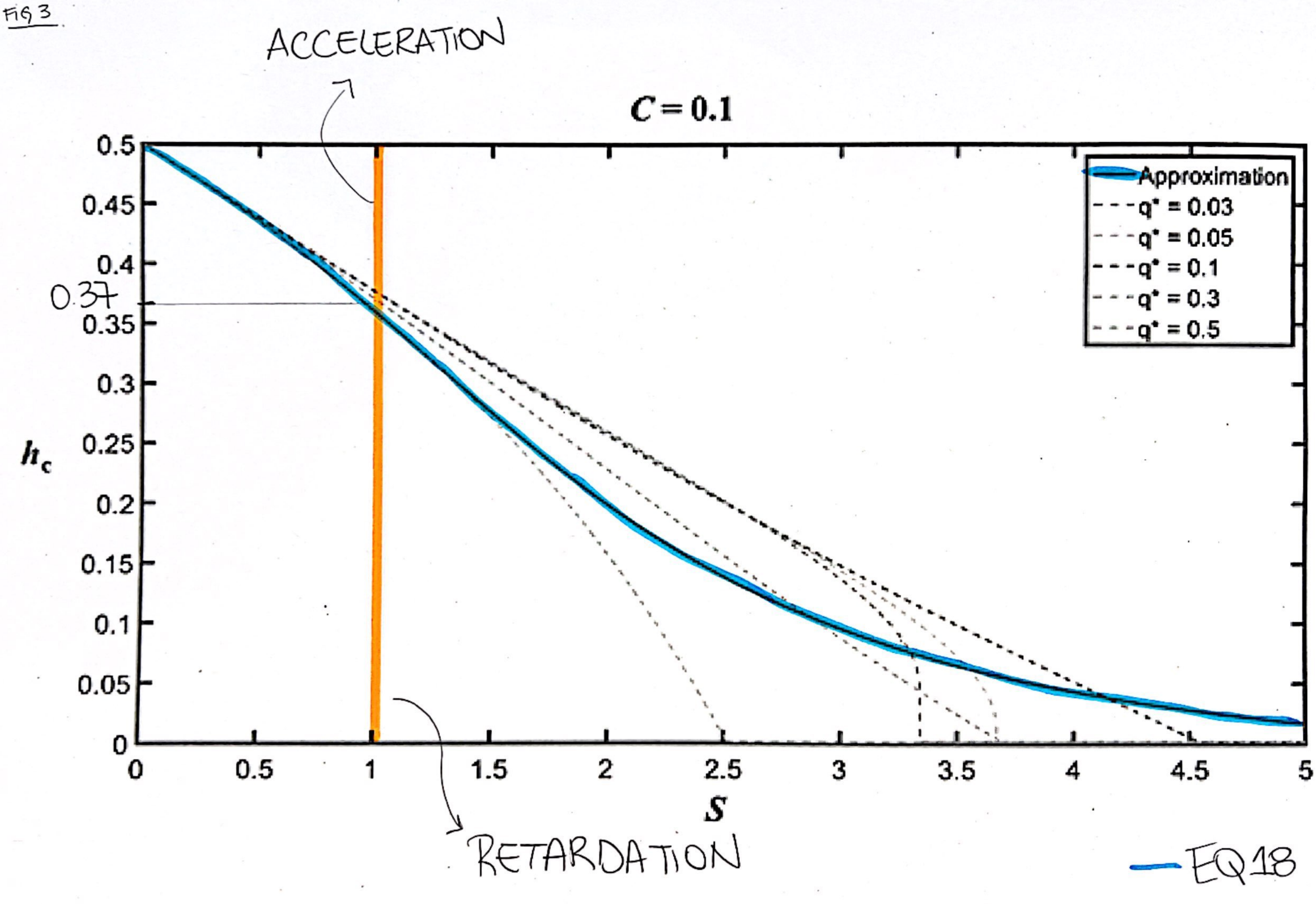
Scanned with CamScanner

FIG2



Scanned with CamScanner

Fig 3


ACCELERATION
C = 0.1
0.37
Approximation
q* = 0.03
q* = 0.05
q* = 0.1
q* = 0.3
q* = 0.5
h_c
S
RETARDATION
EQ18