LECTURE: How Drift Affects the Variance Between Populations

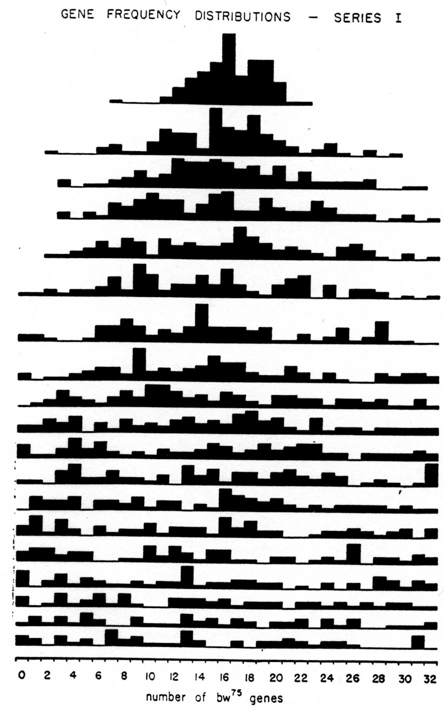


FIG. 6. Gene frequency distributions by generation in series I. Graphical presentation of the data of table 13.

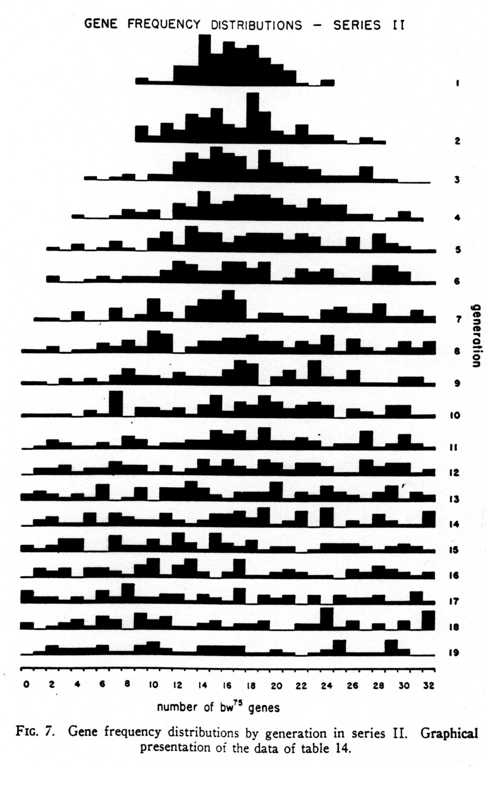


FIG. 7. Gene frequency distributions by generation in series II. Graphical presentation of the data of table 14.

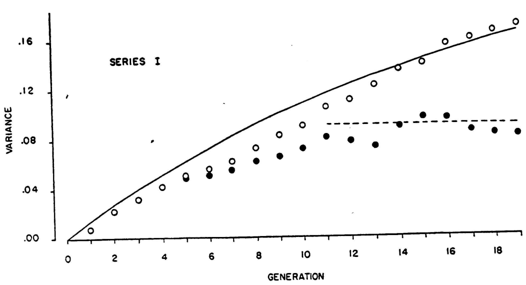


FIG. 12. Theoretical variances of the total frequency distribution by generation including fixed classes and based on a common estimate of $2N_e = 18$ for series I are represented by the smooth curves. Open circles show the observed variance of the distribution including previously fixed classes. Closed circles indicate the observed total variance excluding fixed classes. The asymptote ($= 0.091$) indicates approximately the theoretical maximum value of this variance. All values are on a relative scale.

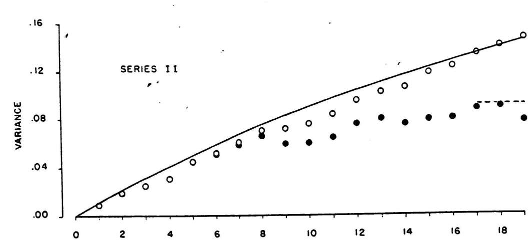


FIG. 13. Theoretical variances of the total frequency distribution by generation including fixed classes and based on a common estimate of $2N_e = 23$ for series II are represented by the smooth curve. Open circles show the observed variance of the distribution including previously fixed classes. Closed circles indicate the observed total variance excluding fixed classes. The asymptote ($= 0.091$) indicates approximately the theoretical maximum value of this variance. All values are on a relative scale.

We need a theory for variation between populations. Let us build one on top of what we have already: a theory for gene diversity within populations.

Notation:

$$
\begin{aligned}
p &= \text{frequency of } A_1 \text{ in generation 0} \\
x_t &= \text{frequency of } A_1 \text{ in generation } t \\
\mathcal{H}_0 &= 2p(1-p), \text{ initial gene diversity} \\
\mathcal{H}_t &= \text{gene diversity in generation } t \\
E[\mathcal{H}_t|x_0] &= \text{expectation of } \mathcal{H}_t \text{ given } x_0 \\
V &= E[x_t^2 - p^2|p], \text{ variance of } x_t
\end{aligned}
$$

We know that

$$E[\mathcal{H}_t|p] = 2p(1-p)\left(1 - \frac{1}{2N}\right)^t$$

since $2p(1-p)$ is the initial gene diversity. We also know that

$$E[\mathcal{H}_t|x_t] = 2x_t(1-x_t)$$

since $2x_t(1-x_t)$ is the diversity in generation t.

These expressions are conditioned on allele frequencies in different generations. To reconcile them, we can write

$$2p(1-p)\left(1 - \frac{1}{2N}\right)^t = E[2x_t(1-x_t)|p]$$

The right-hand side of this expression is

$$
\begin{aligned}
E[2(x_t - x_t^2)|p] &= 2(p - E[x_t^2|p]) \qquad \text{(see B.9)} \\
&= 2[p - (V + p^2)] \qquad \text{(see B.2)} \\
&= 2[p(1-p) - V]
\end{aligned}
$$

Setting this equal to the left-hand side of the previous equation gives

$$V = p(1-p)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right]$$

We usually normalize this expression by dividing both sides by $p(1-p)$. The result is called F_{st}:

$$F_{st} = \frac{V}{p(1-p)} = 1 - \left(1 - \frac{1}{2N}\right)^t$$

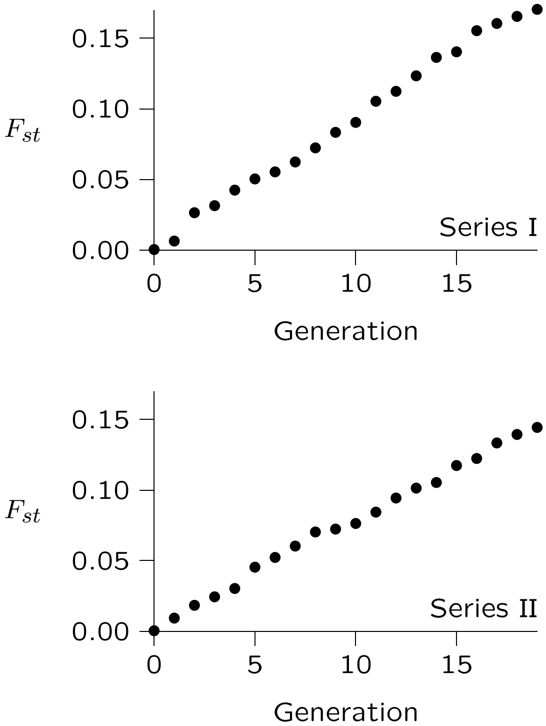
We can estimate F_{st} from data using

$$F_{st} = \frac{V}{p(1-p)}$$

Note that p is an *ancestral* allele frequency. We can predict its value using

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

Data from Buri (1956)



Data from Buri (1956)