Mechanics of allele-frequency change I

(Not-so) random mating and genetic drift

1. Words for what's being copied and transmitted

2. Frequencies of alleles and genotypes

3. Effects of population size and behavior over time (drift in an experiment, and a model)

Three systems of vocabulary

	1	2	3
Position on chromosome	locus	locus	locus
Protein-coding locus	gene	gene	gene
Physical copy of DNA at locus	gene	allele	gene copy
One of several variants at a locus	allele	allele	allele

1 is classical usage, 2 is Gillespie's, and we try to keep to 3.

Illustration of classical usage

> Those organisms (homozygotes) which received like genes, in any pair of corresponding loci, from their two parents, would necessarily hand on genes of this kind to all of their offspring alike; whereas those (heterozygotes) which received from their two parents genes of different kinds. . . (Fisher, 1930, p. 8)

The same sentence in the three systems

Classical	If the *genes* you inherited from mom and dad are different alleles, then you are a heterozygote.
Gillespie	If the *alleles* you inherit from mom and dad are different alleles, then you are a heterozygote.
Us	If the *gene copies* you inherit from mom and dad are different alleles, then you are a heterozygote.

Transferrin genotype frequencies in a baboon troop

G'type	Number of baboons	C	D
CC	80	160	0
CD	15	15	15
DD	5	0	10
Total	100	175	25

Relative frequency

$$\hat{x}_{CC} = 80/100 = 0.80$$
$$\hat{x}_{CD} = 15/100 = 0.15$$
$$\hat{x}_{DD} = 5/100 = 0.05$$
$$\hat{p} = 175/200 = 0.875$$

Note: "hat" indicates values describing sample rather than population. I'll often ignore this distinction.

Alternative calculation of p

$$\begin{aligned} \hat{p} &= \hat{x}_{CC} + \hat{x}_{CD}/2 \\ &= 0.80 + 0.15/2 = 0.875 \end{aligned}$$

The sample allele frequency $\hat{p}$ is an estimate of the population allele frequency p.

The population allele frequency is also the probability that a gene drawn at random from the population is a copy of allele C.

Allele frequency as probability

Suppose there are two alleles, A_1 and A_2, with frequencies p and $1 - p$. What is the probability that a random gene copy is an A_1?

It is just the relative frequency, p, of a allele A_1 within the population.

You can also think of it this way: select a random individual, and from that individual choose a random gene. You end up with A_1 with probability

$$p = P_{11} \times 1 + P_{12} \times \frac{1}{2}$$

where P_{11} and P_{12} are the frequencies of genotypes A_1A_1 and A_1A_2.

Expected genotype frequencies

What is the probability that a random baboon will have genotype CD?

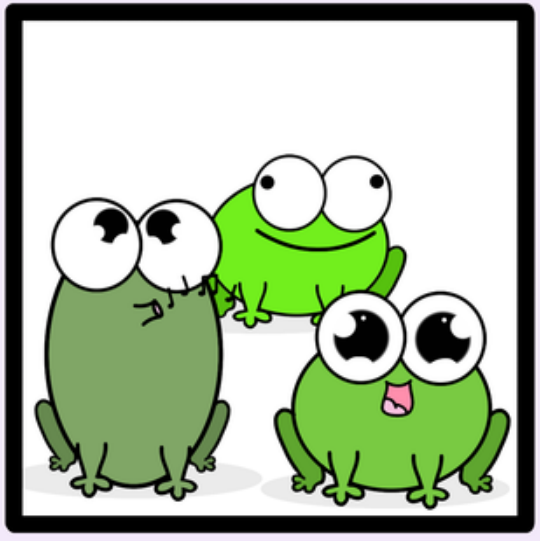
If we know the genotype frequencies, the answer is x_{CD}, the genotype frequency.

But what if we only know the allele frequency?

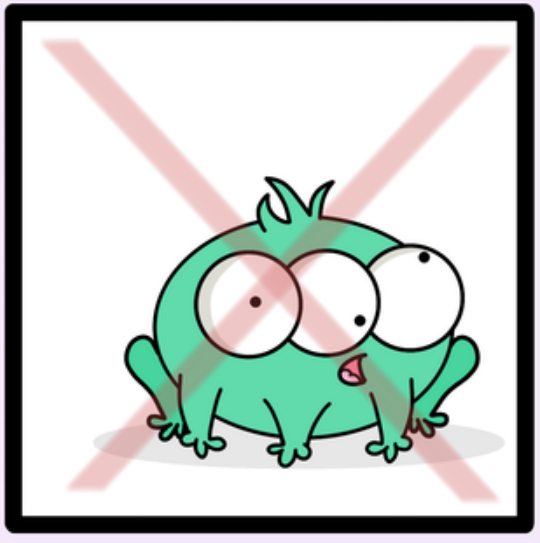
Then the answer depends on characteristics of population. To describe these effects, we need a model.

Assumptions of Hardy-Weinberg Equilibrium

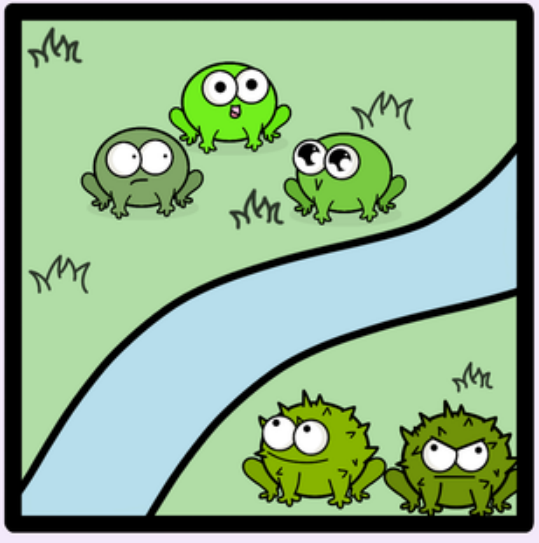
1. No Selection



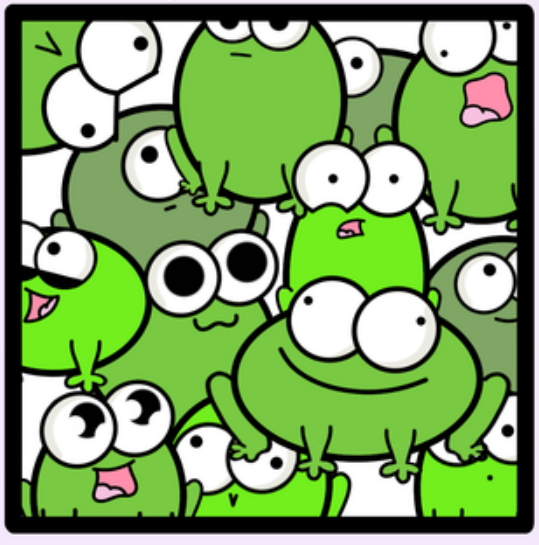
2. No Mutation



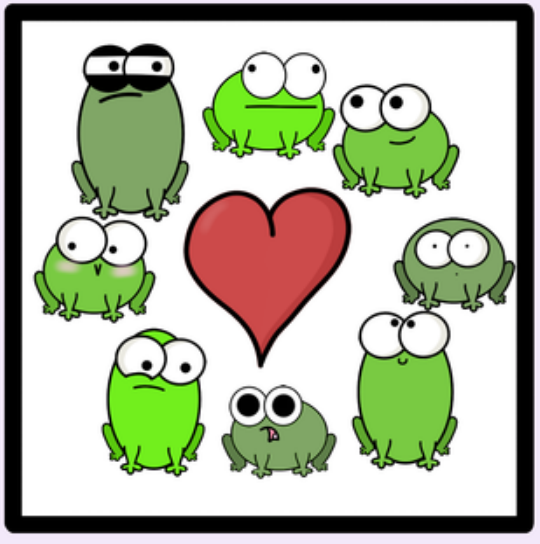
3. No Migration



4. Large Population



5. Random Mating



@AmoebaSisters

Model: random mating, no selection

Event CD can be decomposed as follows:

Gene copy from			
Mom	Dad	Probability	
C	D	$p \times (1-p)$	Why multiply?
D	C	$(1-p) \times p$	Why multiply?
Sum:		$2p(1-p)$	Why add?

Event CC

Gene copy from			
Mom	Dad	Probability	
C	C	$p \times p$	Why multiply?
	Sum:	p^2	

Hardy-Weinberg result

Genotype	Relative frequency
CC	$x_{CC} = p^2$
CD	$x_{CD} = 2pq$
DD	$x_{DD} = q^2$

Where $q = 1 - p$.

- Random mating does not change p.
- Given allele frequency, we can predict genotype frequencies.

This assumes an infinite population with random mating and no selection. Real populations aren't like that, so why should we care about Hardy-Weinberg?

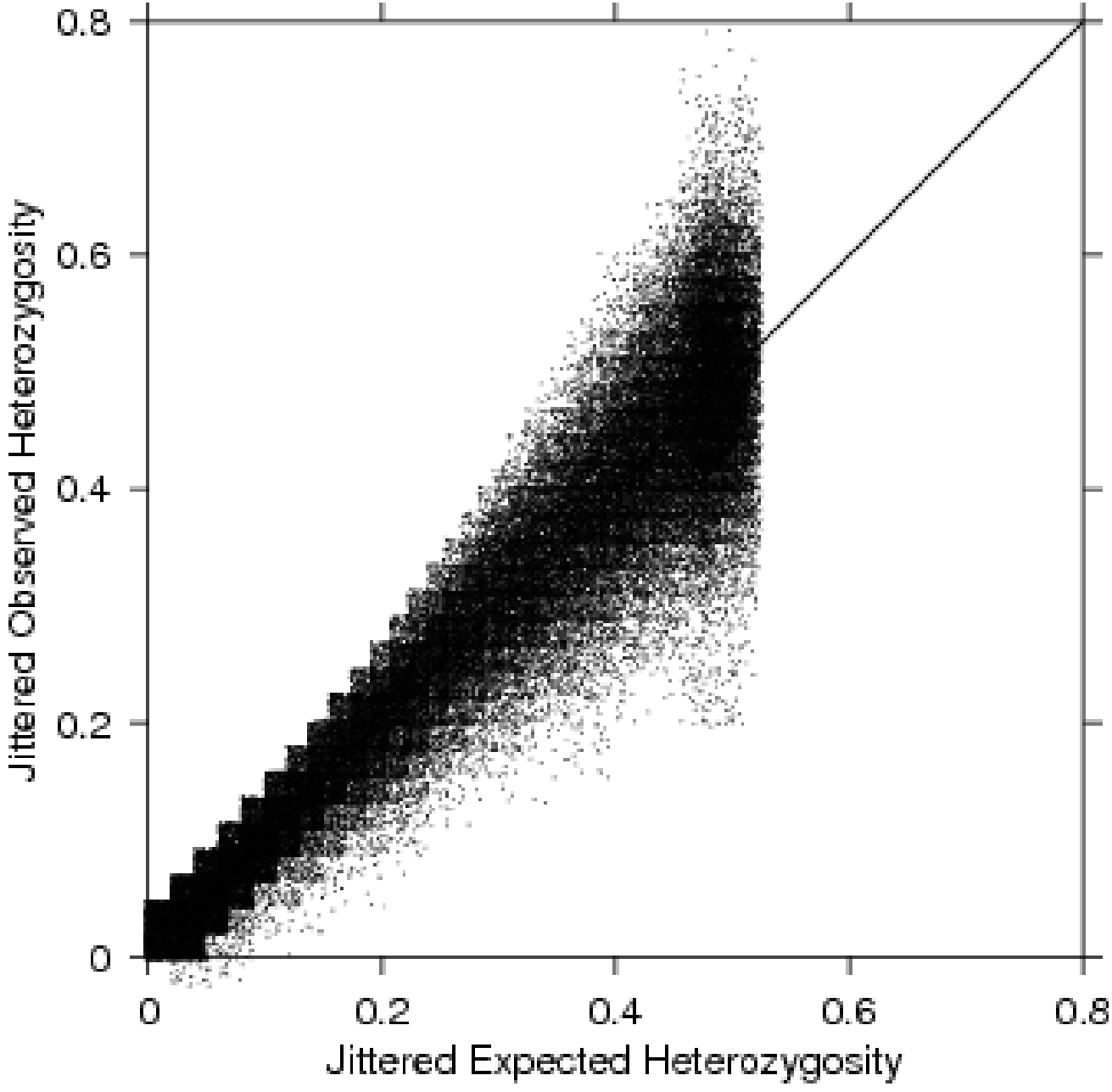
Observed versus expected g'type freqs

	Relative frequency	
Genotype	Observed	Expected
CC	$x_{CC} = 0.80$	$p^2 = 0.77$
CD	$x_{CD} = 0.15$	$2pq = 0.22$
DD	$x_{DD} = 0.05$	$q^2 = 0.02$

Observed: relatative frequency of genotype in data

Expected: Hardy-Weinberg formula

Heterozygosity on human chromosome 1

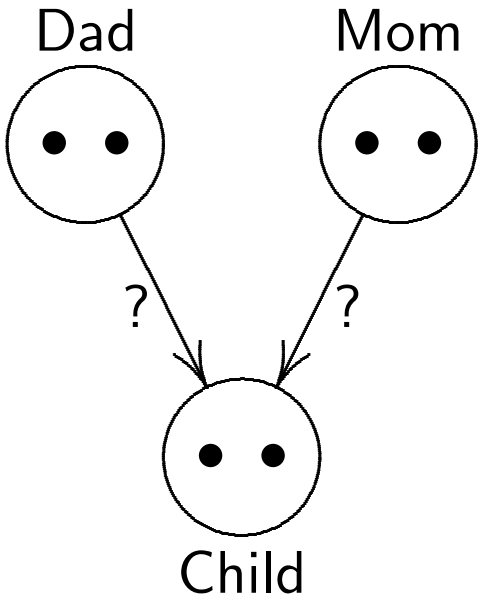


What if males and females have different allele frequencies?

	Genotype frequencies		
Sex	A_1A_1	A_1A_0	A_0A_0
♂	x_{11}	x_{10}	x_{00}
♀	y_{11}	y_{10}	y_{00}

Sex	Allele frequency
♂	$p_m = x_{11} + x_{10}/2$
♀	$p_f = y_{11} + y_{10}/2$

An autosomal locus in a nuclear family



Probabilities that gametes carry A_1

$$\begin{array}{ll} \male & x_{11} + x_{10}/2 = p_m \\ \female & y_{11} + y_{10}/2 = p_f \end{array}$$

Child genotype probabilities

$$\begin{aligned} x'_{11} &= p_m p_f \\ x'_{10} &= p_m(1-p_f) + p_f(1-p_m) \\ x'_{00} &= (1-p_m)(1-p_f) \end{aligned}$$

After one generation of random mating, the sexes have equal allele frequencies at autosomal loci.

$$\begin{aligned} p' &= x'_{11} + x'_{10}/2 \\ &= (p_m + p_f)/2 \end{aligned}$$

Summary

- At equilibrium under random mating, allele frequencies determine genotype frequencies.
- Hermaphrodites reaches equilibrium in 1 generation.
- Autosomal loci in sexual populations reach equilibrium in 2 generations.
- X-linked loci in reach equilibrium only gradually.

Mechanics of allele-frequency change I

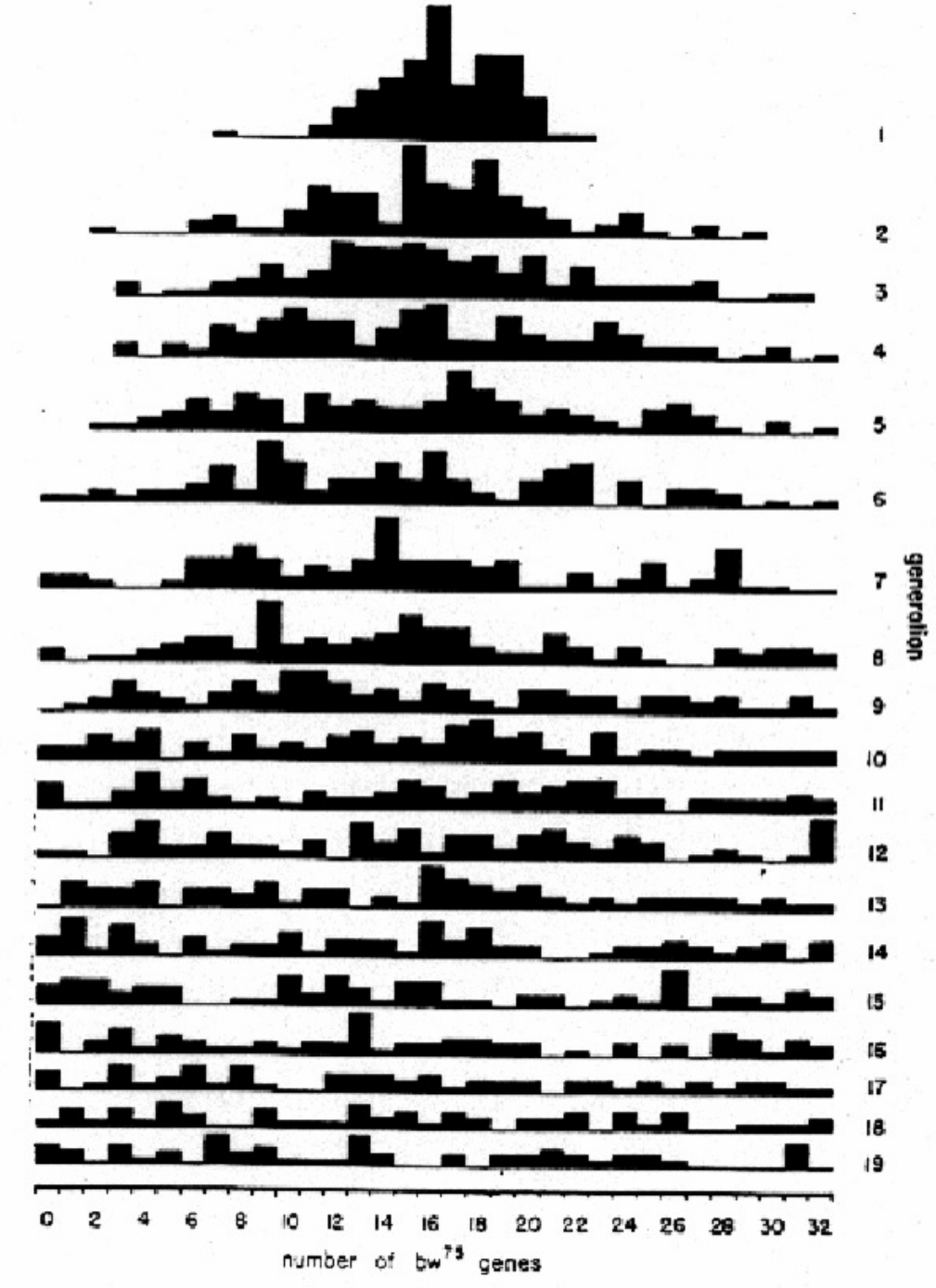
(Not-so) random mating and genetic drift

1. Words for what's being copied and transmitted

2. Frequencies of alleles and genotypes

3. Effects of population size and behavior over time (drift in an experiment, and a model)

We begin with data from an experiment, described by Peter Buri in 1956. (Gene frequency in small populations of mutant *Drosophila*, Evolution, 10:367–402)

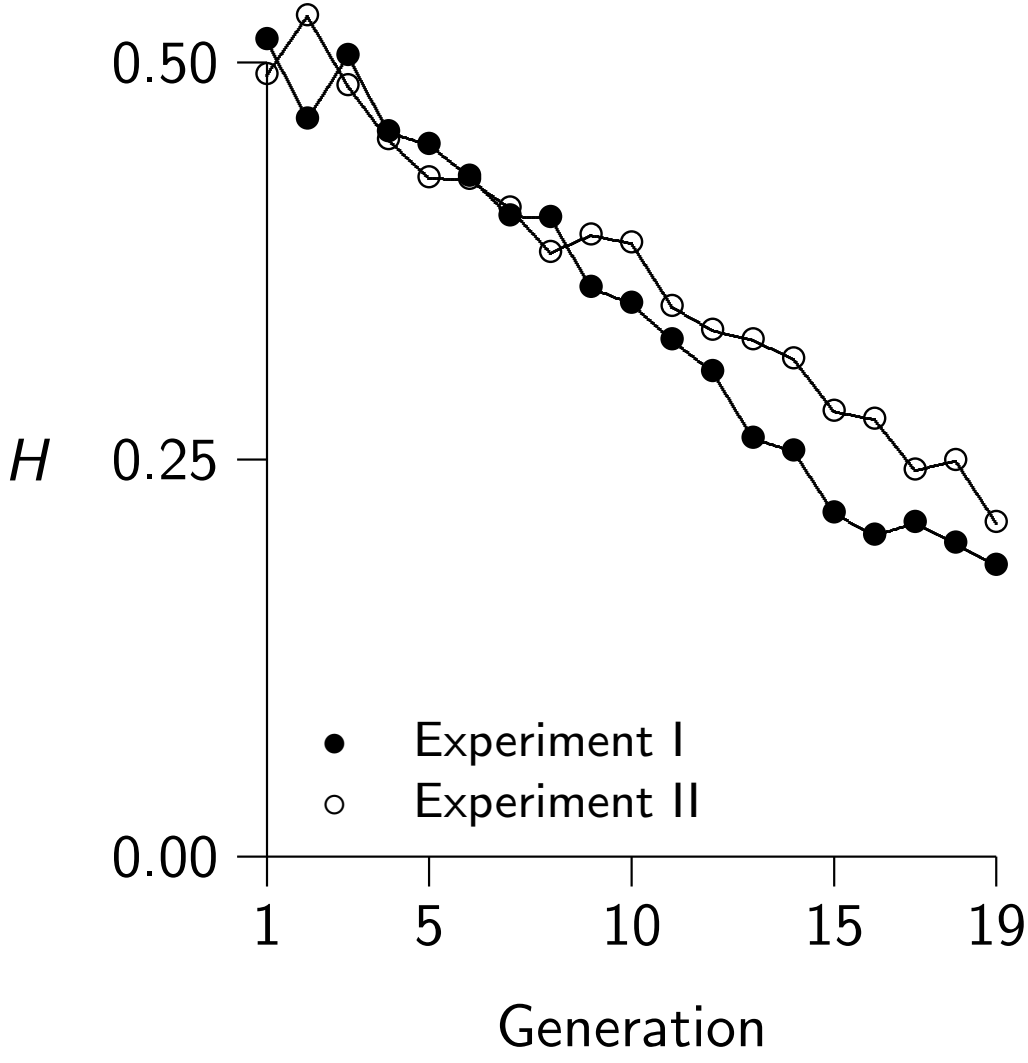


Buri's drift experiment I

- Each generation: 107 bottles, each w/ 8 male & 8 female fruit flies.
- Generation 0: all flies heterozygous.
- Rows show distribution of allele frequency in 19 successive generations.

Peter Buri, 1956

Decay of heterozygosity in Buri's two experiments



- Heterozygosity (H) starts at 0.5
- Declines to about 0.2
- Why?

As heterozygosity declined w/i bottles, the variance among them increased

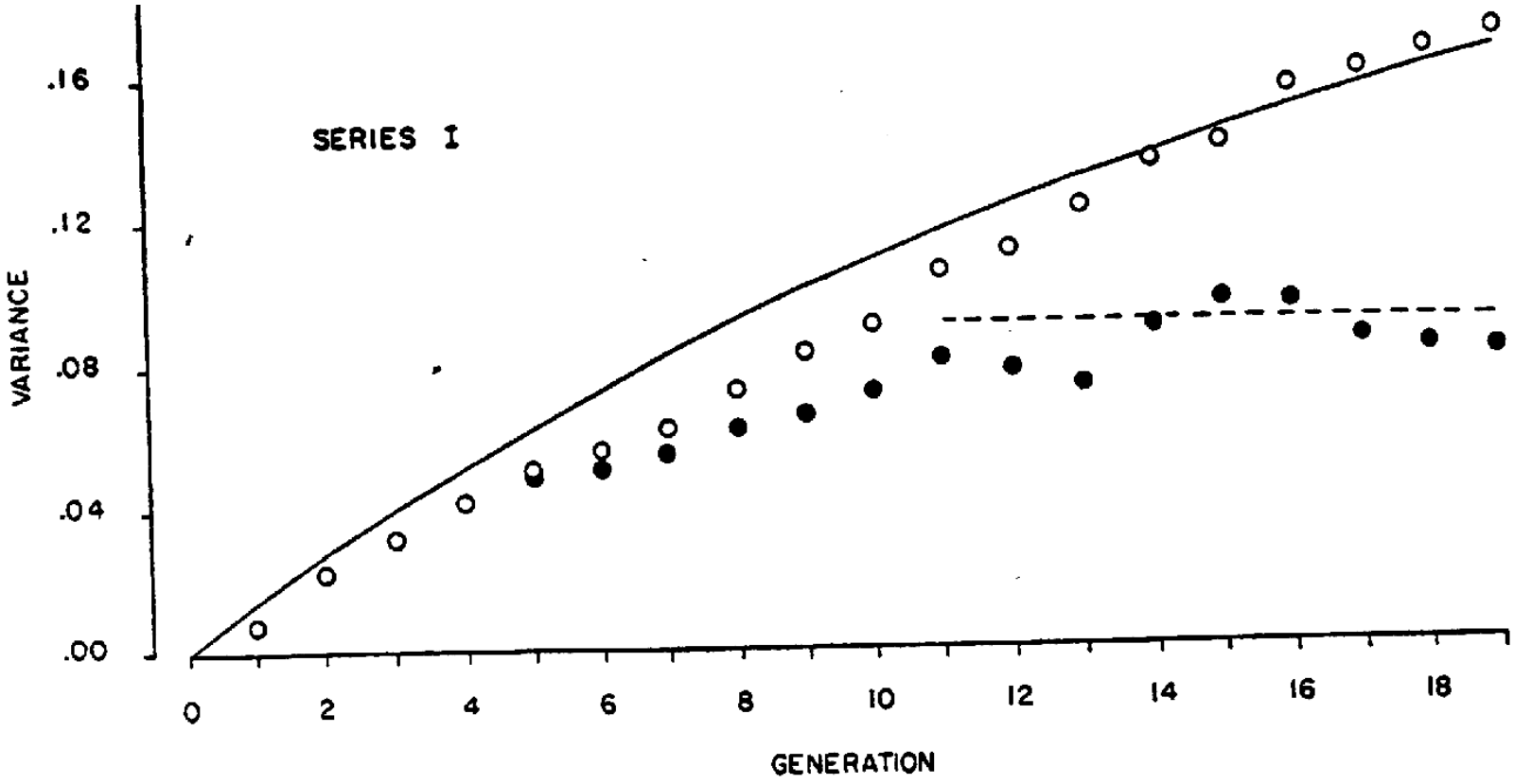
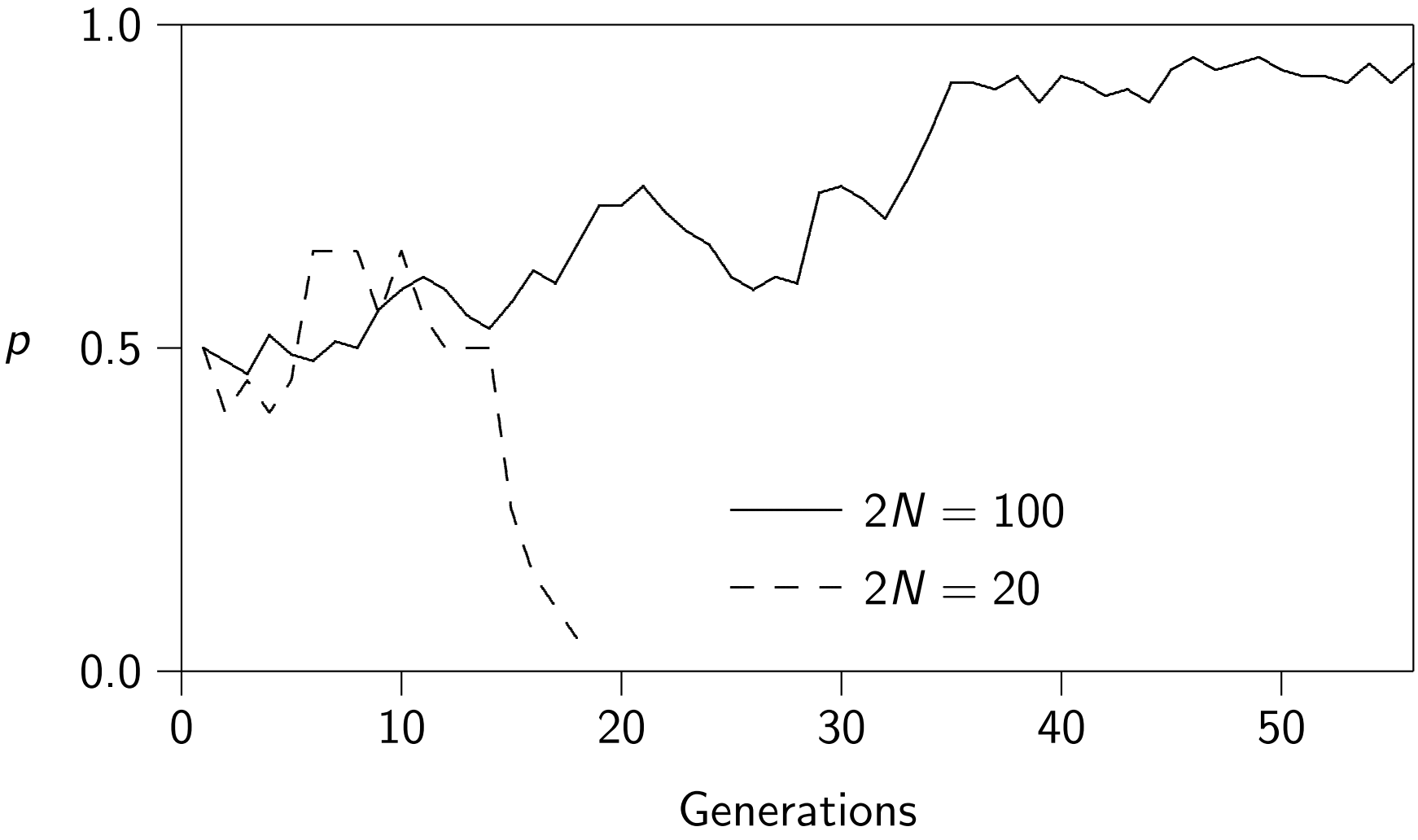


FIG. 12. Theoretical variances of the total frequency distribution by generation including fixed classes and based on a common estimate of $2N_e = 18$ for series I are represented by the smooth curves. Open circles show the observed variance of the distribution including previously fixed classes. Closed 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.

Computer Simulations of Genetic Drift



The Urn Metaphor

Imagine two urns: metaphors for a population in two successive generations. Urn 1 has 50 balls, some red, some white, representing parental gene copies. Urn 2 is empty until urn 1 has "reproduced" as follows:

1. Examine a random ball from urn 1.
2. Put a ball of the same color into urn 2.
3. Replace the ball from urn 1.
4. Repeat until there are 50 balls in urn 2.

Urn 2 is differs from urn 1 because of random sampling: a metaphor for genetic drift.

The urn model behaves a lot like genetic drift in real populations:

1. variation between populations increases
2. variation within populations decreases

Yet real organisms don't reproduce as our urns do. The best urn model is unlikely to be one in which the number of balls matches the number of gene copies.

Decay of Heterozygosity: Notation

$N = \#$ of diploid individuals in population

$2N = \#$ of gene copies in population

$\mathcal{G} =$ Probability that two random gene copies, drawn with replacement from generation t, are copies of the same allele.

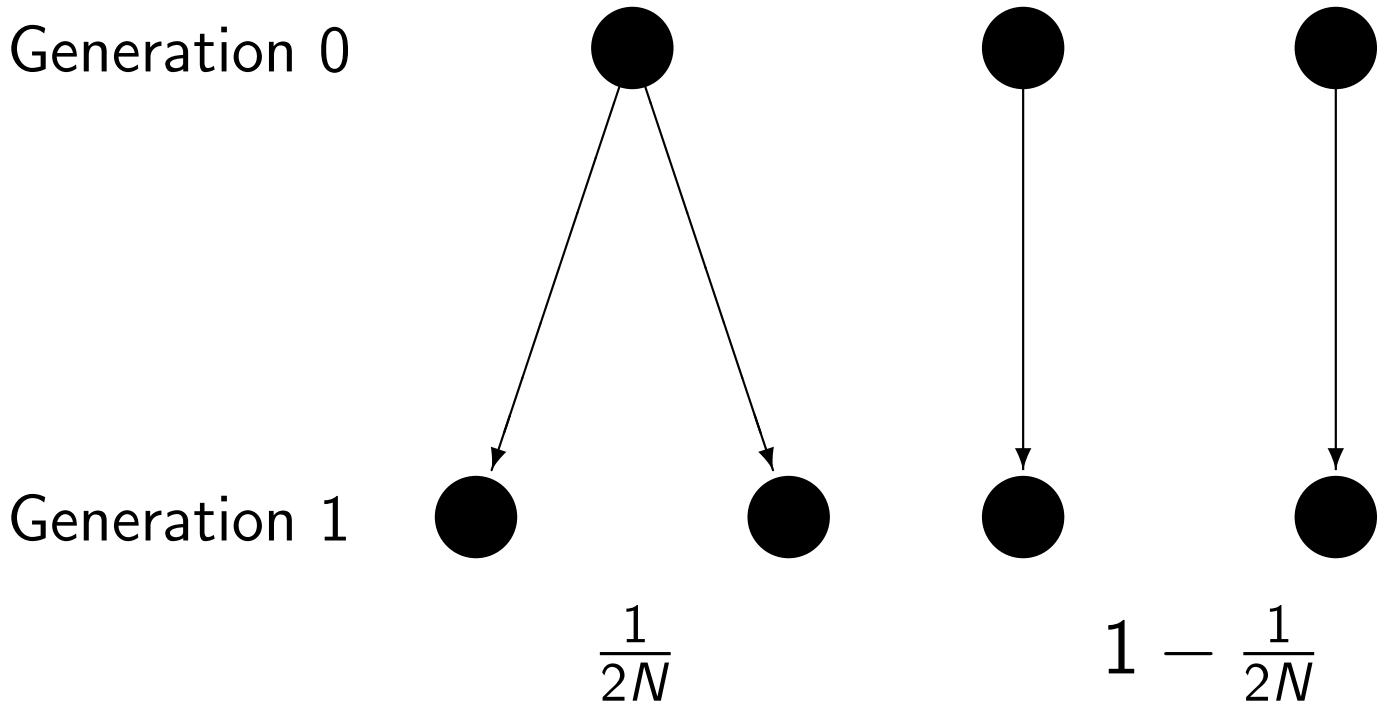
$\mathcal{G}' =$ same thing in the generation $t + 1$.

Decay of Heterozygosity: Logic

Two gene copies may be identical in state either because

1. they are copies of the same parental gene copy, or
2. they are copies of distinct parental gene copies, which happen to be identical in state.

Two gene copies either are or are not copies of the same parental gene copy



Two gene copies are copies of the same parental gene copy with probability $1/2N$, and of distinct parental gene copies with probability $1 - 1/2N$.

Event	Prob
Individual carries 2 copies of same parental gene copy	$1/2N$

Explanation:

1. First draw a random gamete from among those produced by the parental generation. This gamete is equally likely to have been produced by any of the $2N$ parental gene copies.
2. Next draw another gamete at random. There is 1 chance in $2N$ that the second is a copy of the same parental gene copy as the first.

Event	Prob
Individual carries copies of 2 distinct parental gene copies, which are themselves identical.	$(1 - 1/2N)\mathcal{G}$

Explanation:

1. The two random gene copies are copies of distinct parental genes with probability $1 - 1/2N$.
2. These distinct parental gene copies are copies of the same allele with probability $\mathcal{G}$—that is the definition of $\mathcal{G}$.
3. *Both* things are true with probability:

$$\left(1 - \frac{1}{2N}\right)\mathcal{G}$$

In short, the two genes are identical if they are copies either of

1. the same parental gene copy (probability $1/2N$), or of
2. distinct but identical gene copies (probability $(1 - 1/2N)\mathcal{G}$).

Altogether,

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}$$

To see where this goes, it is easier to work with the probability that the two gene copies are copies of *different* alleles, i.e. with the heterozygosity,

$$\begin{aligned} \mathcal{H}' &= 1 - \mathcal{G}' \\ &= \left(1 - \frac{1}{2N}\right)\mathcal{H} \qquad \text{(after some algebra).} \end{aligned}$$

Can you supply the algebra?

The Time-path of Heterozygosity

$$
\begin{aligned}
\mathcal{H}_1 &= \left(1 - \frac{1}{2N}\right)\mathcal{H}_0 \\
\mathcal{H}_2 &= \left(1 - \frac{1}{2N}\right)\mathcal{H}_1 \\
&= \left(1 - \frac{1}{2N}\right)^2\mathcal{H}_0 \\
\mathcal{H}_t &= \left(1 - \frac{1}{2N}\right)^t\mathcal{H}_0
\end{aligned}
$$

where $\mathcal{H}_0$ is the original heterozygosity and $\mathcal{H}_t$ is the heterozygosity in generation t.

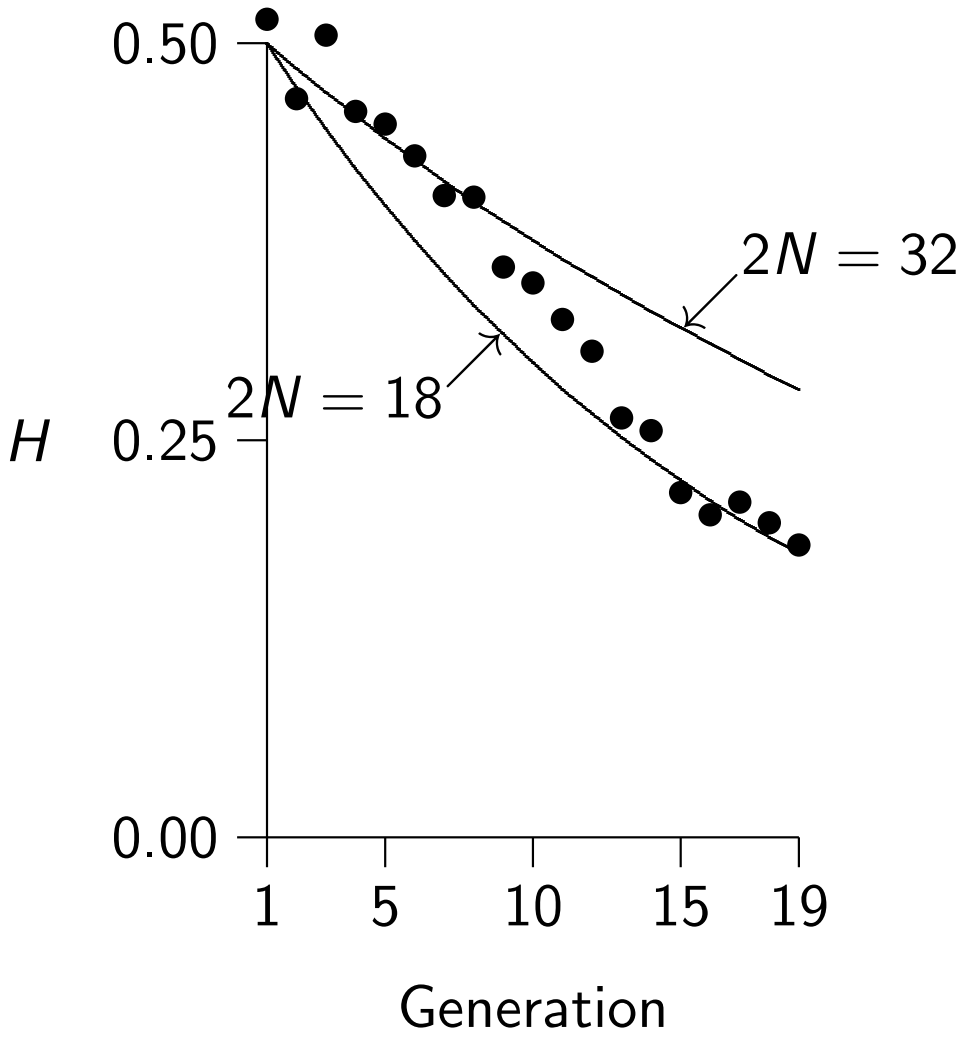
Example

In Peter Buri's experiment, $\mathcal{H}_1 = 1/2$ because half the population were heterozygotes after the first generation of random mating. 18 generations later:

$$\mathcal{H}_{19} = \frac{1}{2}\left(1 - \frac{1}{2N}\right)^{18}$$

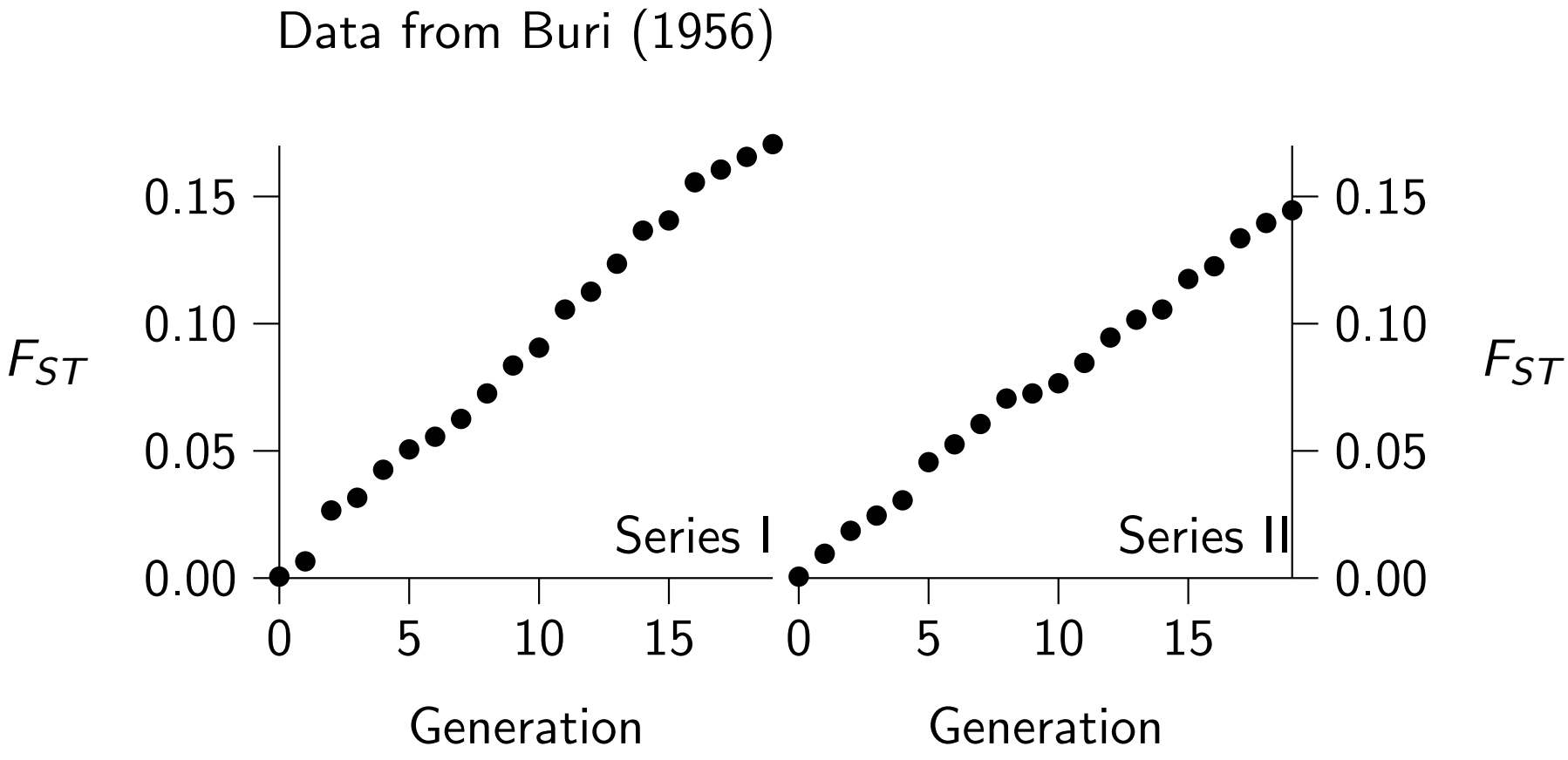
But what is $2N$?

Heterozygosity: Buri's experiment I vs. urn model

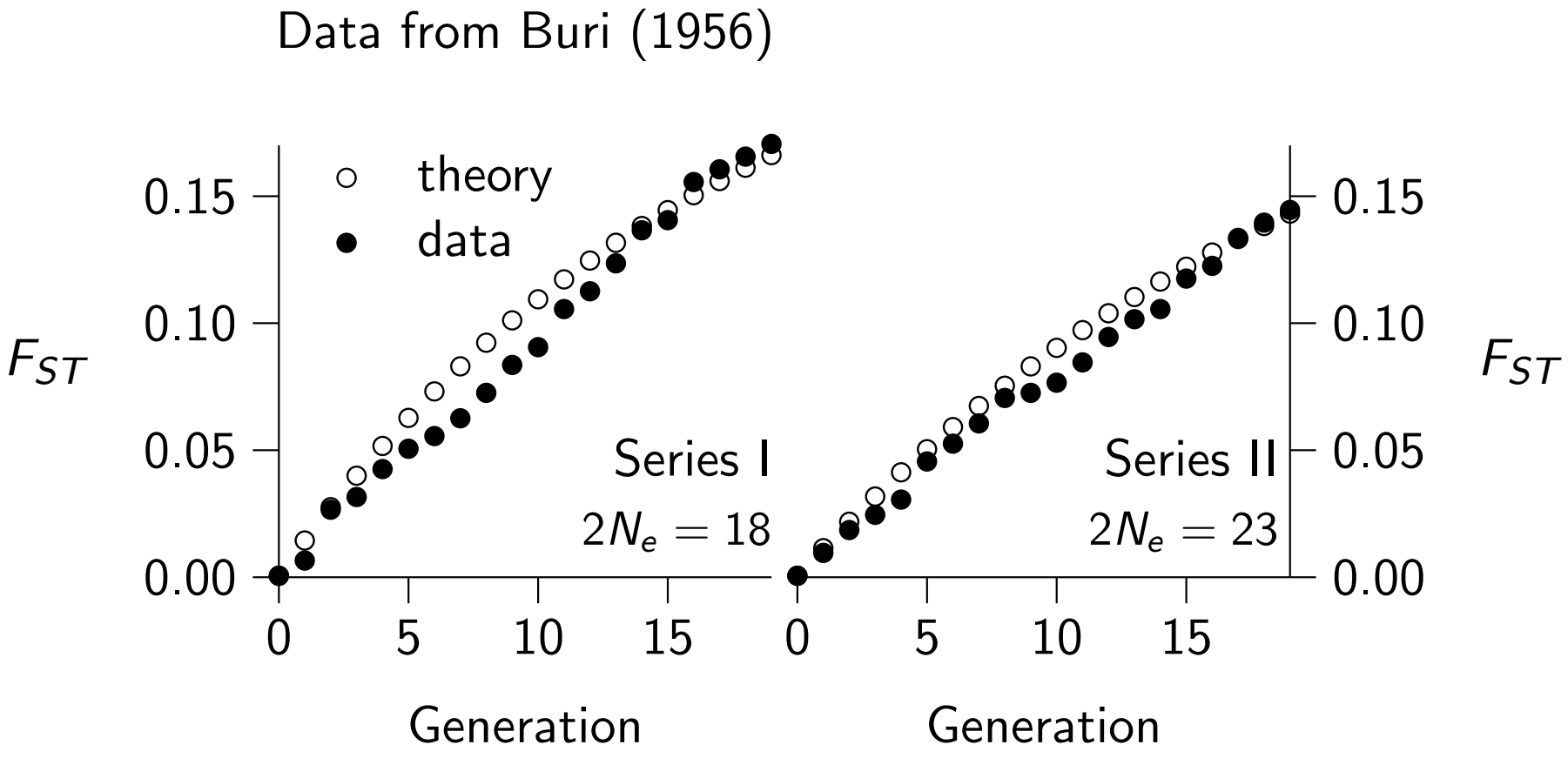


- There were 32 gene copies in each bottle.
- Yet $2N = 32$ provides a poor fit to data.
- Better fit with $2N = 18$.
- 18 is the "*effective population size*"

F_{ST} measures variation among populations



Model fits after setting $N = N_e$



Mechanics of allele-frequency change I (Summary)

1. For sanity, please distinguish genes, gene copies and alleles!

2. Mating isn't truly "random", and doesn't need to be!

3. Unless some other process increases heterozygosity, it will decay exponentially at a rate proportional to 1/Ne.