Alleles and Genotypes in Populations that Mate at Random

Alan R. Rogers

January 7, 2024

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?

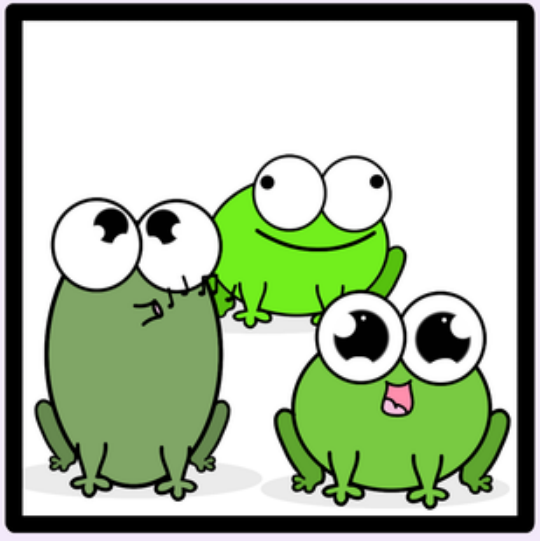
If we know the genotype frequencies, the answer is x_{CD}, the genotype frequency.

But what if we only know the allele frequency?

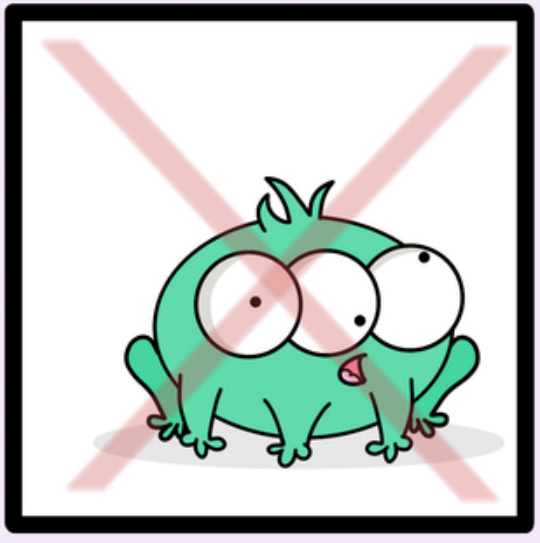
Then the answer depends on characteristics of population. To describe these effects, we need a model.

Assumptions of Hardy-Weinberg Equilibrium

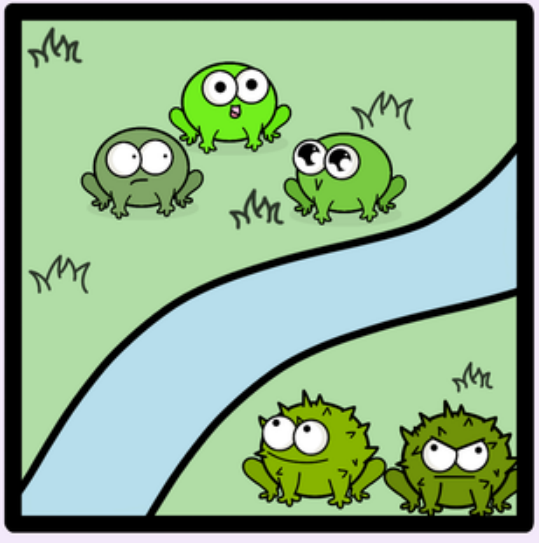
1. No Selection



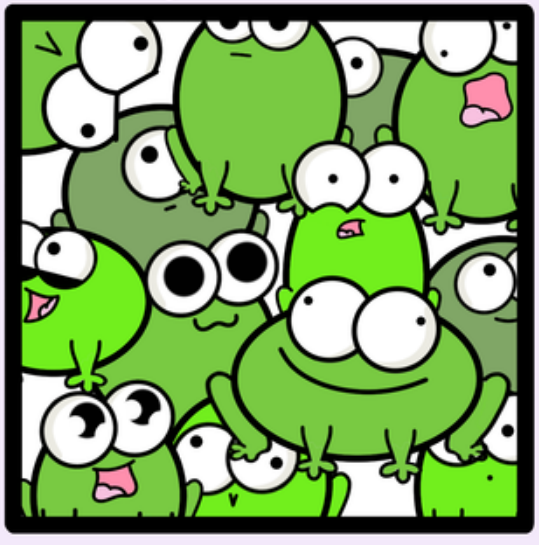
2. No Mutation



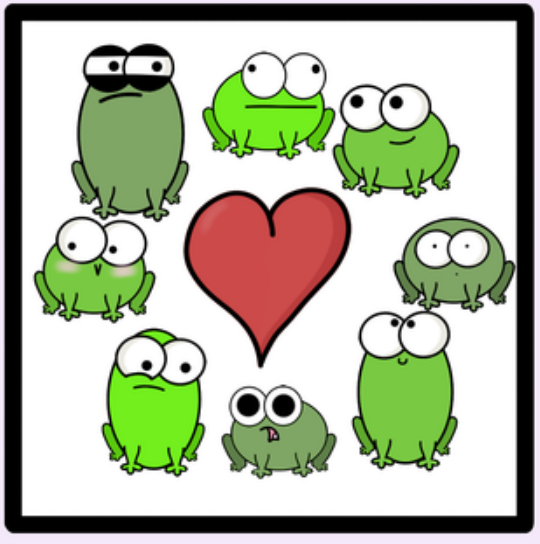
3. No Migration



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?

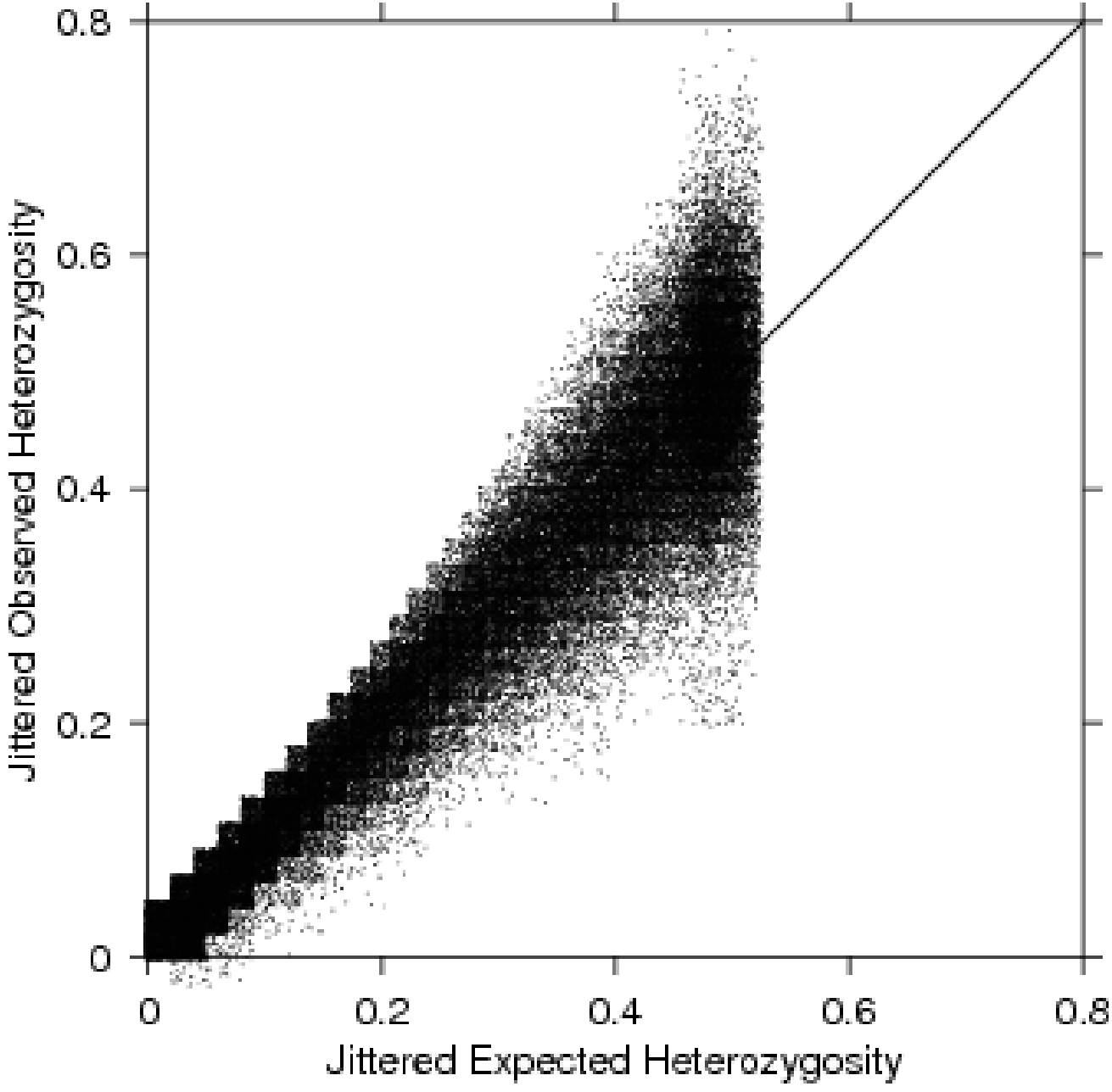
Observed versus expected g'type freqs

	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

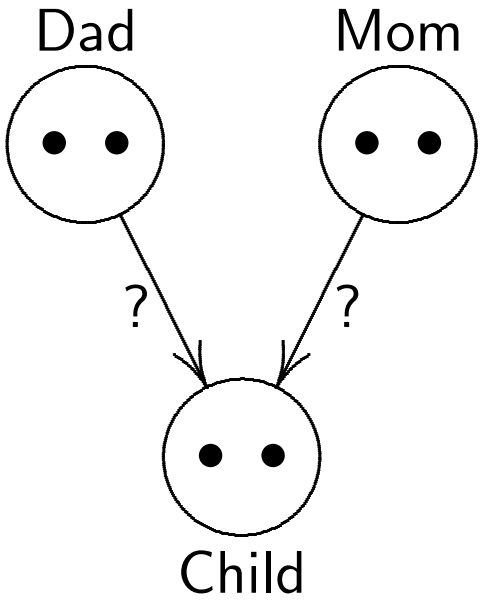


What if males and females have different allele frequencies?

	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.