

Random Mating

Alan R. Rogers

January 22, 2021

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 prefer 3.

Consistency is the hobgoblin of little minds

—Ralph Waldo Emerson

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. . . (R.A. Fisher, 1930, p. 8)

1
 atg.tcg.ttt.act.ttg.acc.aac.aag.aac.gtg.att.ttc.gtt.gcc.ggt.ctg.gga.ggc.att.ggt
 Met.Ser.Phe.Thr.Leu.Thr.Asn.Lys.Asn.Val.Ile.Phe.Val.Ala.Gly.Leu.Gly.Gly.Ile.Gly
 61
 ctg.gac.acc.agc.aag.gag.ctg.ctc.aag.cgc.gat.ctg.aag.aac.ctg.gtg.atc.ctc.gac.cgc
 Leu.Asp.Thr.Ser.Lys.Glu.Leu.Leu.Lys.Arg.Asp.Leu.Lys.Asn.Leu.Val.Ile.Leu.Asp.Arg
 121
 att.gag.aac.ccg.gct.gcc.att.gcc.gag.ctg.aag.gca.atc.aat.cca.aag.gtg.acc.gtc.acc
 Ile.Glu.Asn.Pro.Ala.Ala.Ile.Ala.Glu.Leu.Lys.Ala.Ile.Asn.Pro.Lys.Val.Thr.Val.Thr
 181
 t
 ttc.tac.ccc.tat.gat.gtg.acc.gtg.ccc.att.gcc.gag.acc.acc.aag.ctg.ctg.aag.acc.atc
 Phe.Tyr.Pro.Tyr.Asp.Val.Thr.Val.Pro.Ile.Ala.Glu.Thr.Thr.Lys.Leu.Leu.Lys.Thr.Ile
 241
 ttc.gcc.cag.ctg.aag.acc.gtc.gat.gtc.ctg.atc.aac.gga.gct.ggt.atc.ctg.gac.gat.cac
 Phe.Ala.Gln.Leu.Lys.Thr.Val.Asp.Val.Leu.Ile.Asn.Gly.Ala.Gly.Ile.Leu.Asp.Asp.His
 301
 cag.atc.gag.cgc.acc.att.gcc.gtc.aac.tac.act.ggc.ctg.gtc.aac.acc.acg.acg.gcc.att
 Gln.Ile.Glu.Arg.Thr.Ile.Ala.Val.Asn.Tyr.Thr.Gly.Leu.Val.Asn.Thr.Thr.Thr.Ala.Ile
 361
 t a
 ctg.gac.ttc.tgg.gac.aag.cgc.aag.ggc.ggt.ccc.ggt.ggt.atc.atc.tgc.aac.att.gga.tcc
 Leu.Asp.Phe.Trp.Asp.Lys.Arg.Lys.Gly.Gly.Pro.Gly.Gly.Ile.Ile.Cys.Asn.Ile.Gly.Ser
 421
 a
 gtc.act.gga.ttc.aat.gcc.atc.tac.cag.gtg.ccc.gtc.tac.tcc.ggc.acc.aag.gcc.gcc.gtg
 Val.Thr.Gly.Phe.Asn.Ala.Ile.Tyr.Gln.Val.Pro.Val.Tyr.Ser.Gly.Thr.Lys.Ala.Ala.Val
 481
 a c g t
 gtc.aac.ttc.acc.agc.tcc.ctg.gcg.aaa.ctg.gcc.ccc.att.acc.ggc.gtg.acc.gct.tac.acc
 Val.Asn.Phe.Thr.Ser.Ser.Leu.Ala.Lys.Leu.Ala.Pro.Ile.Thr.Gly.Val.Thr.Ala.Tyr.Thr
 541
 c
 gtg.aac.ccc.ggc.atc.acc.cgc.acc.acc.ctg.gtg.cac.aag.ttc.aac.tcc.tgg.ttg.gat.gtt
 Val.Asn.Pro.Gly.Ile.Thr.Arg.Thr.Thr.Leu.Val.His.Lys.Phe.Asn.Ser.Trp.Leu.Asp.Val
 601
 t c c
 gag.ccc.cag.gtt.gct.gag.aag.ctc.ctg.gct.cat.ccc.acc.cag.cca.tcg.ttg.gcc.tgc.gcc
 Glu.Pro.Gln.Val.Ala.Glu.Lys.Leu.Leu.Ala.His.Pro.Thr.Gln.Pro.Ser.Leu.Ala.Cys.Ala
 661
 a
 gag.aac.ttc.gtc.aag.gct.atc.gag.ctg.aac.cag.aac.gga.gcc.atc.tgg.aaa.ctg.gac.ctg
 Glu.Asn.Phe.Val.Lys.Ala.Ile.Glu.Leu.Asn.Gln.Asn.Gly.Ala.Ile.Trp.Lys.Leu.Asp.Leu
 721
 ggc.acc.ctg.gag.gcc.atc.cag.tgg.acc.aag.cac.tgg.gac.tcc.ggc.atc.
 Gly.Thr.Leu.Glu.Ala.Ile.Gln.Trp.Thr.Lys.His.Trp.Asp.Ser.Gly.Ile.

Alcohol
 dehydrogenase
 coding region in *D.
 melanogaster*

(Gillespie 2004)

Nucleotides &
 amino acids.

Most in 3rd codon
 position.

Only 1 (pos 578)
 changes amino acid.

```

atg.tcg.ttt.act.ttg.acc.aac.aag.aac.gtg.att.ttc.gtt.gcc.ggt.ctg.gga.ggc.att.ggt
.g a. c. c. . . . . . c. . . . . g. . . . . c
.Ala. . . . . . . . . . . . . . . . . . . . .
ctg.gac.acc.agc.aag.gag.ctg.ctc.aag.cgc.gat.ctg.aag.aac.ctg.gtg.atc.ctc.gac.cgc
. . . . . .g . . . t. . . . . . . . . . .
. . . . . .Val. . . . . . . . . . . . . . .
att.gag.aac.ccg.gct.gcc.att.gcc.gag.ctg.aag.gca.atc.aat.cca.aag.gtg.acc.gtc.acc
. . . . . C. . . . . . . . . . . . . . . . . . . . .
ttc.tac.ccc.tat.gat.gtg.acc.gtg.ccc.att.gcc.gag.acc.acc.aag.ctg.ctg.aag.acc.atc
.t. . . . . . . . . . . . . . . . .g . . . c. . . .
. . . . . . . . . . . . . . . .Ser. . . . .
ttc.gcc.cag.ctg.aag.acc.gtc.gat.gtc.ctg.atc.aac.gga.gct.ggt.atc.ctg.gac.gat.cac
.a . . . . . . . . . . . . . . . . . . . . .t
.Lys. .Thr. . . . . . . . . . . . . . . . .Tyr
cag.atc.gag.cgc.acc.att.gcc.gtc.aac.tac.act.ggc.ctg.gtc.aac.acc.acg.acg.gcc.att
. . . . . . . . . . . . . . . . . . . . . . . . .
ctg.gac.ttc.tgg.gac.aag.cgc.aag.ggc.ggt.ccc.ggt.ggt.atc.atc.tgc.aac.att.gga.tcc
. . . . . . . . . . . . . . . . c. t. . . . . . . . .
gtc.act.gga.ttc.aat.gcc.atc.tac.cag.gtg.ccc.gtc.tac.tcc.ggc.aac.aag.gcc.gcc.gtg
g. . . . . . . . . . . . . . . . . . . . .t. . . . .t. . .
. . . . . . . . . . . . . . . . . . . . . . . . .
gtc.aac.ttc.acc.agc.tcc.ctg.gcg.aaa.ctg.gcc.ccc.att.acc.ggc.gtg.acc.gct.tac.acc
. . . . . . . . . . . . . . . . . . . . . c. . . . .t.
gtg.aac.ccc.ggc.atc.acc.cgc.acc.acc.ctg.gtg.cac.aag.ttc.aac.tcc.tgg.ttg.gat.gtt
. . . . . . . . . . . . . . . . . . . . . . . . .c . . .
. . . . . . . . . . . . . . . . . . . . . . . . .
gag.ccc.cag.gtt.gct.gag.aag.ctc.ctg.gct.cat.ccc.acc.cag.cca.tcg.ttg.gcc.tgc.gcc
. . . . .g. c. . . . . . . . . . . . . . . .a c. . . .t . .
. . . . . . . . . . . . . . . . . . . . .Thr. . .Ser. .
gag.aac.ttc.gtc.aag.gct.atc.gaa.ctg.aac.cag.aac.gga.gcc.atc.tgg.aaa.ctg.gac.ctg
. . . . .t. . . .c. . .g. . .g . . .t. . . . . . . . .
. . . . . . . . . . . . . . .Glu. . . . . . . . . . .
ggc.acc.ctg.gag.gcc.atc.cag.tgg.acc.aag.cac.tgg.gac.tcc.ggc.atc.
. . . . . . . . . .a. .g . . . . . . . . . . .
. . . . . . . . . .Ser. . . . . . . . . . .

```

ADH sequence in 2 species

Line 1: DNA sequence in *D. melanogaster*

Line 2: Nucleotides that differ in *D. erectus*

Line 3: Amino acids that differ in *D. erectus*

24 synonymous differences; 10 amino acid differences

Most species differences are neutral.

Transferrin genotype frequencies in a baboon troop

G'type	Number of		
	baboons	<i>C</i>	<i>D</i>
<i>CC</i>	80	160	0
<i>CD</i>	15	15	15
<i>DD</i>	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 ?

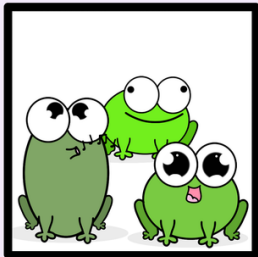
If we know the genotype frequencies, the answer is x_{CD} , the genotype frequency.

But what if we only know the allele frequency?

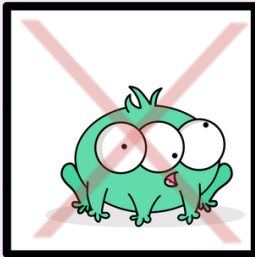
Then the answer depends on characteristics of population. To describe these effects, we need a model.

Assumptions of Hardy-Weinberg Equilibrium

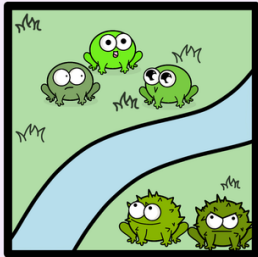
1. No selection



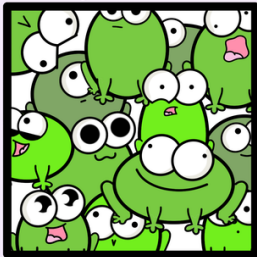
2. No Mutation



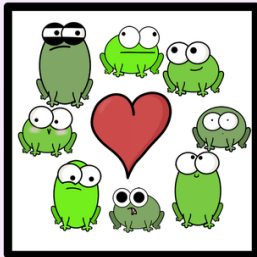
3. No Migration



4. Large Population



5. Random Mating



@AmoebaSisters

Model: random mating, no selection

Event CD can be decomposed as follows:

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

Why multiply?

Why multiply?

Why add?

Event CC

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

Why multiply?

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?

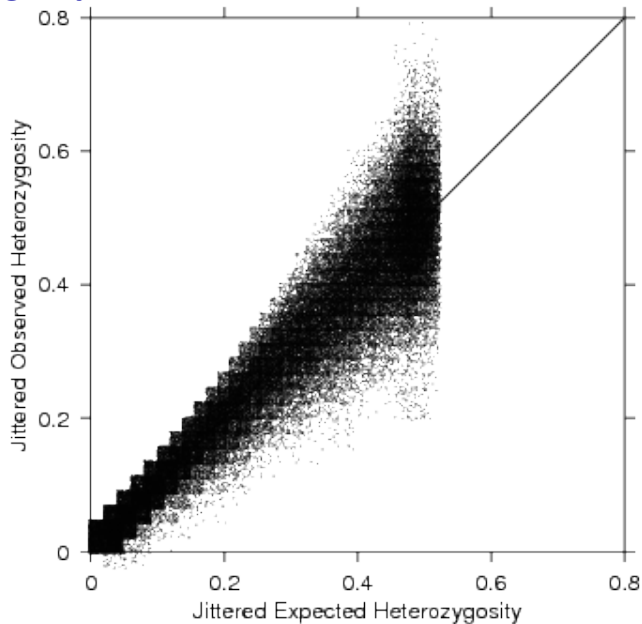
Observed versus expected g'type freqs

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

Observed: relative frequency of genotype in data

Expected: Hardy-Weinberg formula

Heterozygosity on human chromosome 1



What if males and females have different allele frequencies?

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

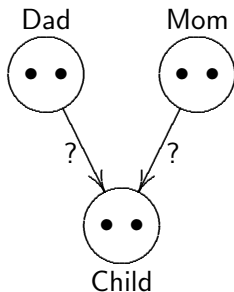
Sex	Allele frequency
♂	$p_m = x_{11} + x_{10}/2$
♀	$p_f = y_{11} + y_{10}/2$

An autosomal locus in a nuclear family

Probabilities that gametes carry A_1

$$\sigma \quad x_{11} + x_{10}/2 = p_m$$

$$\text{♀} \quad y_{11} + y_{10}/2 = p_f$$



Child genotype probabilities

$$x'_{11} = p_m p_f$$

$$x'_{10} = p_m(1 - p_f) + p_f(1 - p_m)$$

$$x'_{00} = (1 - p_m)(1 - p_f)$$

The sexes now have equal allele frequencies.

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

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.