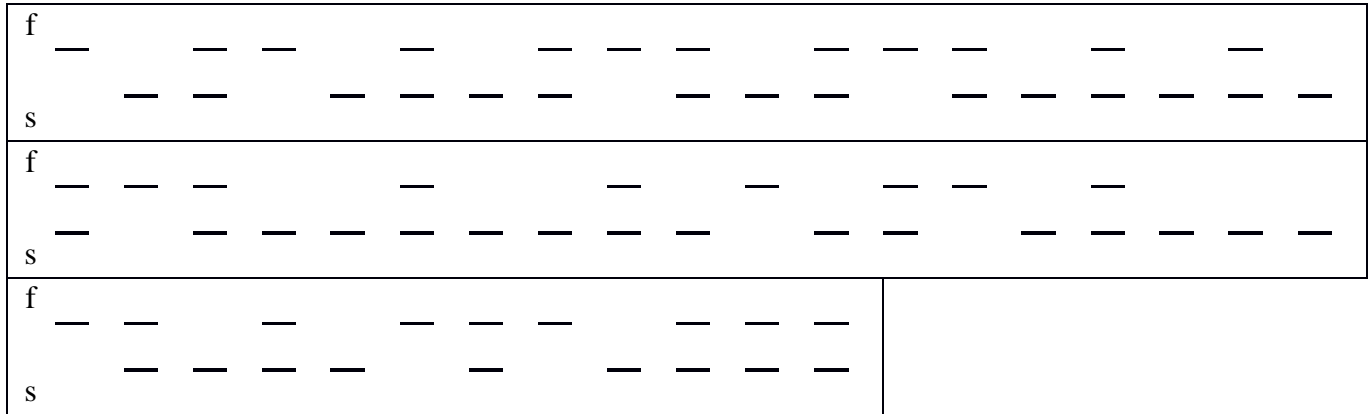


**Part I: Allele frequencies and Hardy-Weinberg equilibrium.**

**Exercise 1**

These data were modified from Meejui *et al.* (2005). They were conducted to quantify genetic variation within stocks of hatchery reared Siamese fighting fish, *Betta splendens*, which is an economically important ornamental fish in Thailand.

*Figure I-1: Diagram of individuals from the population A, locus EST.*



- a- What the diagram figure I-1 represents? **Gel from a run of electrophoresis of allozymes.**
- b- Knowing that f means “fast” and s, “slow”; give the way of the electrolysis on the diagram. **From the bottom to the top.**
- c- How many alleles can we detect on the diagram? **2 alleles.**
- d- What is the frequency of each genotype?  **$ff = 10/50 = 0.2$ ;  $fs = 20/50 = 0.4$ ;  $ss = 20/50 = 0.4$ .**
- e- What is the frequency of each allele?  **$f = ff + \frac{fs}{2} = 0.2 + \frac{0.4}{2} = 0.4$ ;  $s = ss + \frac{fs}{2} = 0.4 + \frac{0.4}{2} = 0.6$**

*Table I-1: Number of expected and observed genotypes.*

Loci	EST			GPI-1					IDHP-1			sMDH-1			MEP			PGM			
	ff	fs	ss	ff	fs	fm	sm	mm	ss	ff	fs	ss	ff	fs	ss	ff	fs	ss	ff	fs	ss
Pop A observed	10	20	20	2	4	6	6	27	5	5	2	43	17	6	37	3	14	33	0	0	50
Pop A expected				1	3	9	13	22	2	1	10	39	7	27	27	2	16	32	0	0	50

- f- The table I-1 gives the alleles frequencies for 5 other loci in the population A. What is the particularity of the locus PGM? **The allele is fixed.**
- g- How can it happen? **Most likely natural selection, genetic drift.**
- h- With n alleles at one locus, how many genotypes exist? **n genotypes homozygote and  $n(n-1)/2$  genotypes heterozygote,**  
**Total amount of genotypes for n alleles at 1 locus:  $\frac{n(n+1)}{2}$**
- i- Under Hardy-Weinberg equilibrium, what would be the genotype frequencies in the next generation, for the 5 loci of the table I-1? And in the 2<sup>nd</sup> generation?

e.g. GPI-1:

$$f = ff + \frac{fs}{2} + \frac{fm}{2} = 2 + \frac{4}{2} + \frac{4}{2} = 7; \frac{7}{(2 + 4 + 6 + 6 + 27 + 5)} = 0.14$$

$$s = ss + \frac{fs}{2} + \frac{sm}{2} = 5 + \frac{4}{2} + \frac{6}{2} = 10; \frac{10}{50} = 0.20$$

$$m = mm + \frac{sm}{2} + \frac{fm}{2} = 27 + \frac{6}{2} + \frac{6}{2} = 33; \frac{33}{50} = 0.66$$

So under Hardy-Weinberg equilibrium,

Allele:		Sperm			
		f	s	m	
Frequency:		p	q	r	
Egg	f	p	ff = p <sup>2</sup>	fs = pq	fm = pr
	s	q	fs = pq	ss = q <sup>2</sup>	sm = qr
	m	r	fm = pr	sm = qr	mm = r <sup>2</sup>

frequency of ff = p<sup>2</sup> = (0.14)<sup>2</sup> = 0.0196; n = 0.0196 x 50 = 0.98 ~1 ind.

frequency of fs = 2pq = 2 x 0.14 x 0.20 = 0.056; n = 0.056 x 50 = 2.8 ~3 ind.

frequency of fm = 2pr = 2 x 0.14 x 0.66 = 0.1848; n = 0.1848 x 50 = 9.24 ~9 ind.

frequency of ss = q<sup>2</sup> = (0.20)<sup>2</sup> = 0.04; n = 0.04 x 50 = 2 ind.

frequency of sm = 2qr = 2 x 0.20 x 0.66 = 0.264; n = 0.264 x 50 = 13.2 ~13 ind.

frequency of mm = r<sup>2</sup> = (0.66)<sup>2</sup> = 0.4356; n = 0.4356 x 50 = 21.78 ~22 ind.

j- What are the amounts of each genotype for each locus? (fill in the table I-1)

k- Does population is at HWE for the locus IDHP-1? for sMDH-1? for MEP? for PGM? for GPI-1?

$$\chi^2 = \sum \frac{(\text{nbr of genot.obs.} - \text{nbr of genot.expected})^2}{\text{nbr of genot.expected}}; \text{d.f.} = (\text{nb genotypes}) - (\text{nb alleles})$$

**GPI-1:**

$$\chi^2 = \frac{(2 - 1)^2}{1} + \frac{(4 - 3)^2}{3} + \frac{(6 - 9)^2}{9} + \frac{(6 - 13)^2}{13} + \frac{(27 - 22)^2}{22} + \frac{(5 - 2)^2}{2}$$

$\chi^2 = 1 + \frac{1}{3} + 1 + 3.769 + 1.136 + \frac{9}{2} = 11.738$ ; d.f. = 3;  $p = 0.039$ ; we can conclude that the observed genotype frequencies at this locus deviate significantly from the HWE.

**IDHP-1:**

$\chi^2 = 22.810$ ; d.f. = 1;  $p < 0.005$ ; we can conclude that the observed genotype frequencies at this locus deviate significantly from the HWE.

**sMDH-1:**

$\chi^2 = 34.323$ ; d.f. = 1;  $p < 0.005$ ; we can conclude that the observed genotype frequencies at this locus deviate significantly from the HWE.

**MEP:**

$\chi^2 = 0.781$ ; d.f. = 1;  $p < 0.677$ ; we can conclude that the observed genotype frequencies at this locus don't deviate significantly from the HWE.

**PGM:**

We can conclude that the observed genotype frequencies at this locus don't deviate significantly from the HWE because the values are identical.

**Exercise 2**

In their study, published in Nature (1964), Spencer *et al.* found three allelic variants (A, B, C) of the red cell acid phosphatase enzyme in a sample of 178 English people. All genotypes were distinguishable by electrophoresis, and the frequencies in the sample were:

Genotype	AA	AB	BB	AC	BC	CC
Frequency (%) observed	9.6	48.3	34.3	2.8	5	0

a- What are the allele frequencies in the sample?

A:  $0.096 + \frac{1}{2}(0.483 + 0.028) = 0.3515$

B:  $0.343 + \frac{1}{2}(0.483 + 0.05) = 0.6095$

C:  $0 + \frac{1}{2}(0.028 + 0.05) = 0.039$

b- Why were no CC individuals found?

Because the expectation is well below 1.

### Exercise 3

*Sine oculis* (so) and Cinnabar (cn) are two autosomal recessive alleles in *Drosophila melanogaster*. They are very closely linked and can be treated as if they were alleles at one locus. The “heterozygote”, so/cn, is wild-type and is distinguishable from both homozygotes; (so/so has no eyes; cn/cn has white eyes if the stock is made homozygous for another eye color mutant, brown, bw). In a class experiment, 4 males and 4 females of an so/so stock were put in a vial together with 16 males and 16 females from a cn/cn stock and allowed to mate. There were 20 such vials. The total count of progeny, classified by genotype, was as follows:

so/so	so/cn	cn/cn
135	359	947

a- How do these numbers differ from Hardy-Weinberg expectations?

The allele frequencies in the parents are known and should be used to calculate the expectations.  $N = 4 + 4 + 16 + 16 = 40$ . The frequency of so was  $(4 + 4) / 40 = 0.2$ , the frequency of cn was  $(16 + 16) / 40 = 0.8$ . The Hardy-Weinberg genotype frequencies in the progeny are:  $0.2^2 = 0.04$ ;  $2 * 0.2 * 0.8 = 0.32$ ;  $0.8^2 = 0.64$ . Multiplying these by the total number counted,  $135 + 359 + 947 = 1441$ , gives the expected numbers (to the nearest whole number) as so/so ~ 58 ind.; so/cn ~ 461 ind.; cn/cn ~ 922 ind. The discrepancy is highly significant, with  $\chi^2 = 102.2 + 22.6 + 0.7 = 125.5$ . This  $\chi^2$  has two degrees of freedom because the observed numbers of progeny were not used to estimate the allele frequency, the only constraint being that the expected number must add to the observed total.

b- Suggest a reason for the discrepancy.

There was an excess of homozygotes and a corresponding deficiency of heterozygotes. Possible reason: assortative mating due to some of the female parents having mated with their own stock males before being put in the vials.

Suppose that the allele frequency in these drosophila cultures is 0.5. This is done by putting 10 males and 10 females of each stock in each vial. The supply of so/so females ran out and only 4 were left for the last vial. So, to preserve the intended allele frequency and numbers of parents, this vial was made up as follows: 16 ♂ + 4 ♀ of so/so with 4 ♂ + 16 ♀ of cn/cn. The student who got this vial was a bit surprised by what you found.

c- What genotype frequencies would you expect in the progeny?

Hardy-Weinberg frequencies are not expected because the allele frequency was different in the male and female parents. The allele frequency of so in males was  $16 / 20 = 0.8$  and in females  $4 / 20$ .

			male	
			so	cn
			$p = 16/20$	$q = 4/20$
female	so	$p = 4/20$	$0.2 * 0.8 = 0.16$	$0.2 * 0.2 = 0.04$

	cn	q = 16/20	0.8*0.8= 0.64	0.8*0.2= 0.16
--	----	-----------	---------------	---------------

Then in the progeny, so/so = 0.16 = 6.4 ind.; so/cn = 0.68 = 27.2 ind.; cn/cn = 0.16 = 6.4 ind.

d- How do these numbers differ from Hardy Weinberg expectations?

With an allele frequency of 0.5 the Hardy-Weinberg expectations are 0.25, 0.5, 0.25. If HWE: so/so = 0.25\*40 = 10; so/cn= 0.5\*40 = 20; cn/cn = 0.25\*40 = 10;  $\chi^2 = 5.184$  and d.f. = 1  $\rightarrow$  p-value < 0.05  $\rightarrow$  we can conclude that the observed genotype frequencies at this locus deviate significantly from HWE.

e- Suggest a reason for the discrepancy.

There is an excess of heterozygotes resulting from the unequal gene frequencies in male and female parents.

#### Exercise 4

In a sample of 1617 Spanish Basques, the numbers of A, B, O, and AB blood types observed were 724, 110, 763, and 20, respectively. These blood groups are due to three alleles,  $I^A$ ,  $I^B$ , and  $I^O$ , with  $I^A I^A$  and  $I^A I^O$  having blood group A;  $I^B I^B$  and  $I^B I^O$  having blood group B.  $I^A I^B$  having blood group AB: and  $I^O I^O$  having blood group O. The best estimates of allele frequency in the Basque sample are 26.61% for  $I^A$ , 4.11% for  $I^B$ , and 69.28% for  $I^O$ .

a- Calculate the expected numbers of the four blood group phenotypes.

A = 724 ind.; B = 110 ind.; O = 763 ind.; AB = 20 ind.; Total N = 1617 ind.

$I^A = 0.2661$ ;  $I^B = 0.0411$ ;  $I^O = 0.6928$ .

Genotype expected:

Blood type A:  $I^A I^A + I^A I^O = p_A^2 + 2 * p_A * p_O = 0.0708 + 0.3687 = 0.4395$  (\*1617 ~ 711 ind.)

Blood type B:  $I^B I^B + I^B I^O = p_B^2 + 2 * p_B * p_O = 0.0586$  (\*1617 ~ 95 ind.)

Blood type O:  $I^O I^O = p_O^2 = 0.4799$  (\*1617 ~ 776 ind.)

Blood type AB:  $I^A I^B = 2 * p_A * p_B = 0.0218$  (\*1617 ~ 35 ind.)

b- Is the population at the Hardy-Weinberg equilibrium for this locus?

$\chi^2 = 0.24+2.37+6.43+0.23 = 9.27$ ; d.f. = 4-3 = 1; p-value < 0.005: we can conclude that the observed genotype frequencies at this locus deviate significantly from the HWE.