Supplemental Material

CBE—Life Sciences Education

Avena *et al*.

SUPPLEMENTAL MATERIALS

Table S1. GLMM comparisons of students' and experts' process use in log odds

	GLMM coefficient estimate (se)		
Process	**incorrect student**	**correct student**	**correct expert**
Notice	-1.03 (0.15)	-0.23 (0.15)	-0.15 (0.37)
Identify similarity	-5.26 (0.68)	0.47 (0.31)	-1.59 (1.27)
Identify concept	-7.10 (0.83)	0.07 (0.36)	2.84 (0.82)
Recall	-1.30 (0.13)	0.05 (0.14)	1.09 (0.28)
Plan	-1.7 (0.14)	0.14 (0.15)	1.84 (0.31)
Check	-3.67 (0.26)	0.50 (0.22)	3.51 (0.42)
Assess difficulty	-4.15 (0.45)	-0.64 (0.27)	1.05 (0.53)
Use information	1.13 (0.11)	-0.21 (0.13)	1.50 (0.35)
Integrate	-1.87 (0.13)	0.45 (0.14)	0.90 (0.27)
Draw	-2.11 (0.19)	-0.39 (0.17)	0.49 (0.39)
Calculate	-0.16 (0.09)	-0.30 (0.11)	-0.23 (0.23)
Reason	2.42 (0.22)	0.96 (0.19)	1.21 (0.49)
Eliminate	-3.53 (0.23)	1.16 (0.22)	1.60 (0.35)
Claim	7.72 (0.92)	0.51 (0.45)	-0.35 (1.07)
Misinterpret	-8.41 (1.13)	-2.80 (1.01)	-1.12 (2.19)
Clarify	-1.53 (0.48)	1.11 (0.25)	6.19 (1.09)
State the process	-5.19 (0.88)	-0.47 (0.31)	1.40 (0.63)
Restate	-3.54 (0.26)	0.24 (0.25)	0.04 (0.49)

The coefficient in the incorrect student column is the intercept in each model, as it is the baseline group for the Expert/Student Answer Status categorical variable. The coefficients for the other two groups in the categorical variable are given in the correct student and correct expert columns. The estimates in column 5 of Table 4 are the inverse logit of the intercepts, and in columns 6 and 7 of Table 4 are the inverse logit of the intercept plus the coefficients for correct student and correct expert answer, respectively.

Table S2. GLMM comparisons of students' process use by content area in log odds

	GLMM coefficient estimate (se)			
Process	**Probability**	**Recombination**	**Nondisjunction**	**Gel/ Pedigree**
Notice	-0.70 (0.18)	-0.35 (0.18)	-0.83 (0.20)	-0.79 (0.20)
Recall	-1.49 (0.17)	1.10 (0.18)	-0.31 (0.20)	-1.44 (0.24)
Plan	-0.93 (0.17)	-1.13 (0.20)	-1.11 (0.21)	-0.67 (0.21)
Check	-3.83 (0.33)	-0.15 (0.32)	0.75 (0.30)	1.16 (0.31)
Assess difficulty	-4.54 (0.48)	-0.02 (0.37)	0.49 (0.36)	0.20 (0.39)
Use information	3.24 (0.25)	-1.37 (0.26)	-4.02 (0.28)	-2.42 (0.26)
Integrate	-1.53 (0.16)	-1.28 (0.22)	-0.73 (0.21)	1.27 (0.19)
Draw	-1.84 (0.23)	-1.31 (0.24)	0.14 (0.22)	-1.76 (0.28)
Reason	3.67 (0.31)	-0.90 (0.30)	-1.10 (0.30)	-0.22 (0.33)

The coefficient in the Probability column is the intercept in each model, as it is the baseline group for the Content Area categorical variable. The coefficients for the other three groups in the categorical variable are given in columns 3 through 5. The estimates in column 6 of Table 5 are the inverse logit of the intercepts, and in columns 7, 8, and 9 are the inverse logit of the intercept plus the coefficients for Recombination, Nondisjunction, and Gel/Pedigree, respectively.

Table S3. For each content area, representative model of a combination of processes that are associated with a correct student answer

	Probability	Recombination	Nondisjunction	Gel/Pedigree
Intercept	-2.54	-2.70	0.87	0.28
Notice				
Plan		0.12		
Recall		1.09		
Check				
Assess difficulty		-0.14		
Use information	0.75	0.99		
Integrate				
Draw	-0.05	-1.39	0.35	
Calculate	2.14	1.15	NA	
Reason	0.39	0.51	0.03	0.65
Eliminate		NA	0.97	

A GLMM model with a lasso penalty was used for variable selection. A representative model predicting answer correctness with a moderate penalty parameter (lambda=25) is shown. The model coefficient shows the positive (green) or negative (red) association with correctness as log odds. Coefficients are additive. Processes with coefficients of 0 in this model are shaded in gray. NA indicates not assessed due to absence of that process in a certain content area.

PRACTICE ASSIGNMENT INSTRUCTIONS AND EXAMPLE

Instructions: In this assignment, please document step-by-step how you are solving each of the problems as you do them. Include the steps you are taking and why, what you are thinking about as you solve the problem, and how you reached your final answer. Please number each of these steps.

Example of how to Document your Problem Solving:

Question: A point mutation in the DMD gene of Jared's DNA is found that changes the nucleotide A to a U and leads to a premature stop codon. What effect will this mutation have on mRNA size?

Documented Problem Solving Answer Example:

1. Read the problem.
2. Know that I am thinking about transcription because the question asks about mRNA.
3. Re-read the problem and note that this mutation is a nonsense mutation because it is a substitution (change in 1 nucleotide) that leads to a stop codon.
4. I know I need to think about the effect of the premature stop codon on length of the mRNA.
5. I draw a strand of DNA and write in the stop codon TAA in the middle so that I can visualize the mutation in the gene.
6. I draw how transcription will occur with an arrow making new mRNA 5' to 3', and I draw the new mRNA, which I realize will include the premature stop codon because I remember that translation will be affected by the codon, NOT transcription.
7. I now link my drawing of transcription to the concept of size of the mRNA and realize that the mRNA for this specific gene will be the same length with or without this specific mutation.
8. FINAL CONCLUSION: This mutation will NOT have any effect on mRNA size.

PRACTICE ASSIGNMENT QUESTIONS

Recombination questions

1. The aldose gene is 10 map units from the fructose gene. (A = dominant allele, can produce aldose; a = recessive allele, cannot produce aldose; F = dominant allele, can produce fructose; f = recessive allele, cannot produce fructose). You cross worms with the chromosomes shown below. What is the probability that offspring will be able to produce only aldose (and not fructose)?

$$\frac{A\ f}{a\ F} \times \frac{a\ f}{a\ f}$$

2. Two genes, Q/q and R/r, are 30 map units apart on a non-sex chromosome and are both inherited in an autosomal recessive manner, with q leading to PKU and r leading to Pendred syndrome.

Michele is the following genotype for two linked genes:

$$\frac{Q\ r}{q\ R}$$

Her husband is the following genotype:

$$\frac{q\ r}{q\ r}$$

What is the chance their child will be a carrier for PKU and have Pendred syndrome?

3. Kumar and Arya are about to have a child. They are worried about two traits that run in the family—fancomi anemia (inherited in an autosomal recessive manner) and polydactyly (inherited in an autosomal dominant manner). The two genes are 14 map units apart. Kumar has fancomi anemia (ff) but does not have polydactyly (dd). Arya is a carrier for fancomi anemia (Ff) and does have polydactyly (Dd).

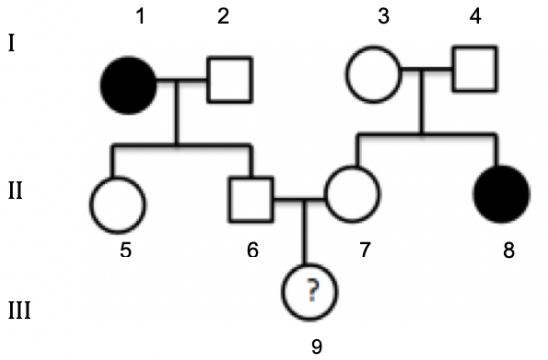
Kumar's genotype is as follows: $\frac{f\ d}{f\ d}$

Arya's genotype is: $\frac{F\ D}{f\ d}$

What is the chance that the child will have both Fancomi Anemia and polydactyly?

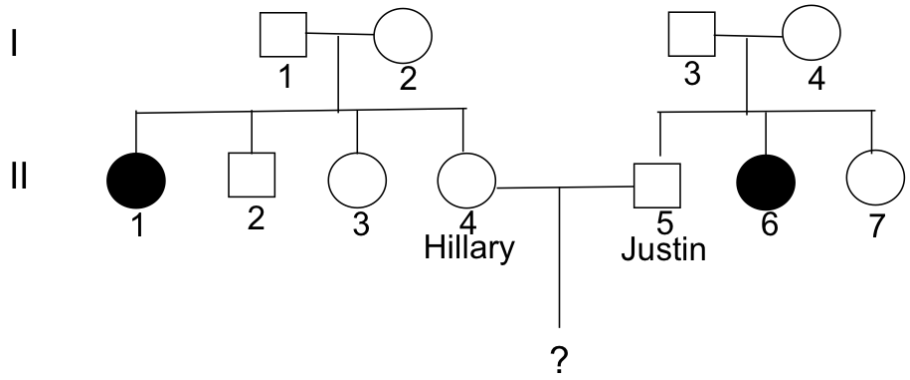
Probability questions

1. Normal giraffes have long necks. Below is a pedigree of a family of giraffes that has a history of an autosomal recessive disorder that causes giraffes to have short necks. What is the probability that the female giraffe with the question mark in the third generation will have the short neck phenotype?

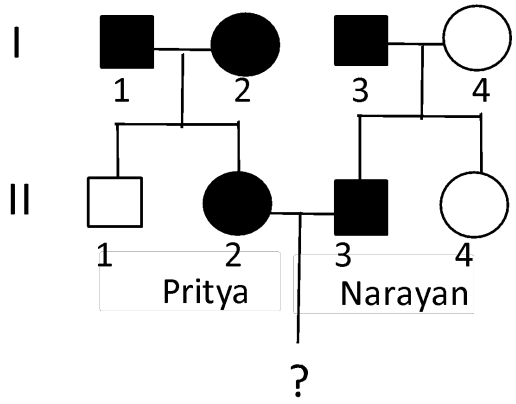


2. Wilson's disease is an inherited disorder in which there is too much copper in the body's tissue. It is inherited in an autosomal recessive manner.

Below are pedigrees from two families where some members of a family have Wilson's disease. Neither Hillary nor Justin have Wilson's disease. If they have a child, what is the chance this child will have Wilson's disease?



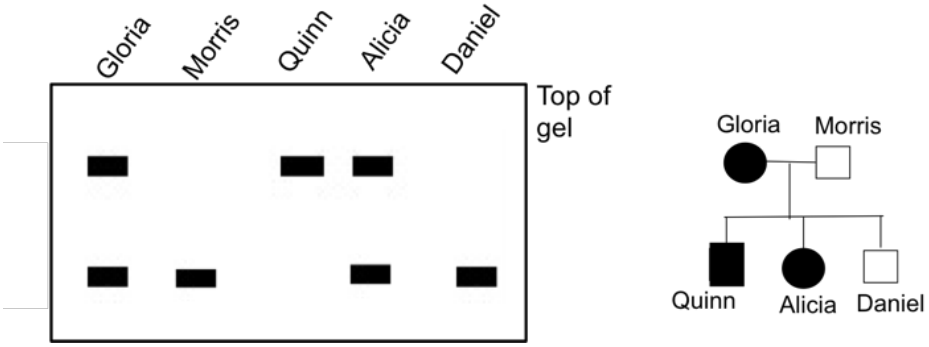
3. The family below has the autosomal dominant trait of dimples. In this family, I-1 and I-2 are known to be heterozygous. It is not known whether any other individuals in this family are heterozygotes. If Narayan and Pritya have a child, what is the chance it will NOT have dimples?



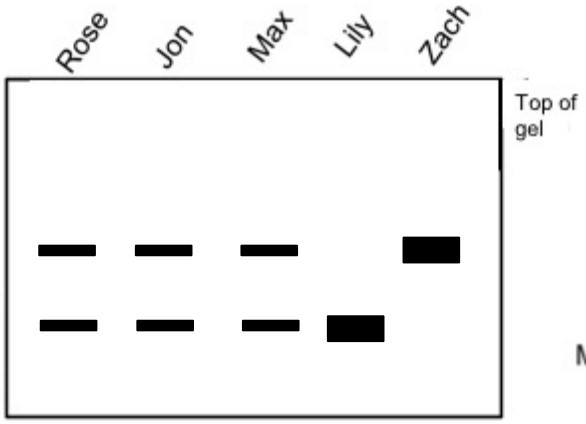
Gel/pedigree questions

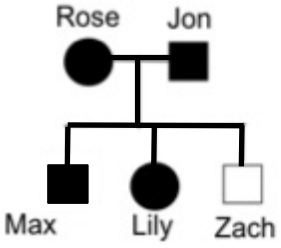
(Note: Students were taught in class that the thickness of a band on a gel indicates relative amount of DNA.)

1. The Clarke family shown below has a genetically inherited disease caused by a 300 bp insertion mutation in an exon of a single gene. DNA for this gene was amplified from each member of the family and run on the gel shown below. Using the pedigree and the gel, what is the most likely mode of inheritance for this disease?

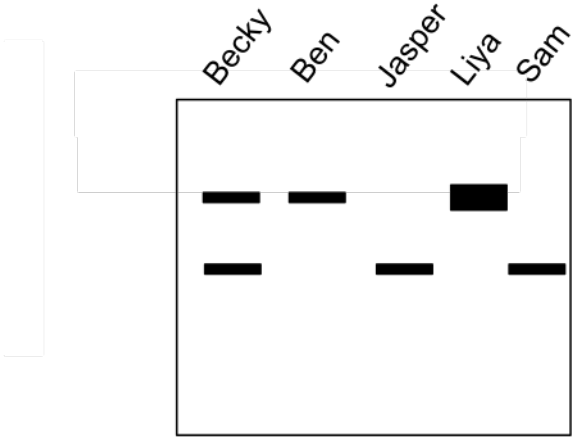


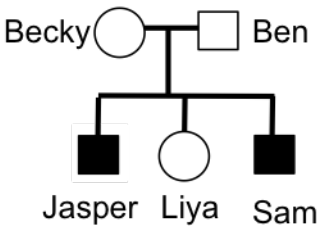
2. The Bustamante family shown below has a genetically inherited disease caused by a 117 bp deletion mutation in an exon of a single gene. DNA for this gene was amplified from each member of the family and run on the gel shown below. Using the pedigree and the gel, what is the most likely mode of inheritance for this disease?





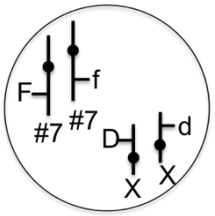
3. The Hale family has a genetically inherited trait called ocular albinism (lack of pigment in the eye) caused by a 516 bp deletion in an exon of a single gene. DNA for this gene was amplified from each member of the family and run on the gel shown below. Using the pedigree and the gel, what is the most likely mode of inheritance for this disease?



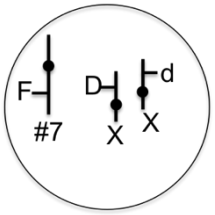


Nondisjunction questions

1. Jill is a carrier for the autosomal recessive disease cystic fibrosis (Ff on chromosome 7). She is also a carrier for the X-linked trait muscular dystrophy ($X^D X^d$). These chromosomes (unreplicated) are shown below in one of Jill's normal cells.

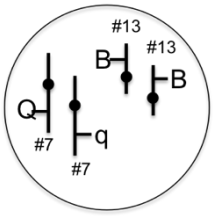


For this problem, assume no crossing over occurs during meiosis. Jill produces an egg with the genotype F $X^D X^d$, as shown below.

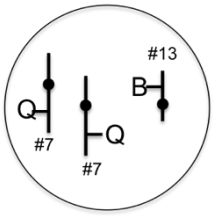


Just considering the steps of meiosis, what has occurred to produce this egg? Specify the phase of meiosis that was affected, if any.

2. Daryl is genotype Qq for a gene on chromosome 7. He is genotype BB for a gene on chromosome 13. These chromosomes (unreplicated) are shown below in one of Daryl's normal cells.

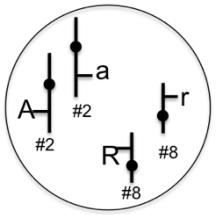


For this problem, assume no crossing over occurs during meiosis. Daryl produces a sperm with genotype BQQ, as shown below.

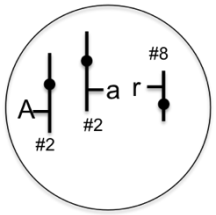


Just considering the steps of meiosis, what has occurred to produce this sperm? Specify the phase of meiosis that was affected, if any.

3. Maria is a carrier for two disorders, with genotype Aa on chromosome 2 and genotype Rr on chromosome 8. These chromosomes (unreplicated) are shown below in one of Maria's normal cells.



For this problem, assume no crossing over occurs during meiosis. Maria produces an egg with genotype Aar, as shown below.



Just considering the steps of meiosis, what has occurred to produce this egg? Specify the phase of meiosis that was affected, if any.