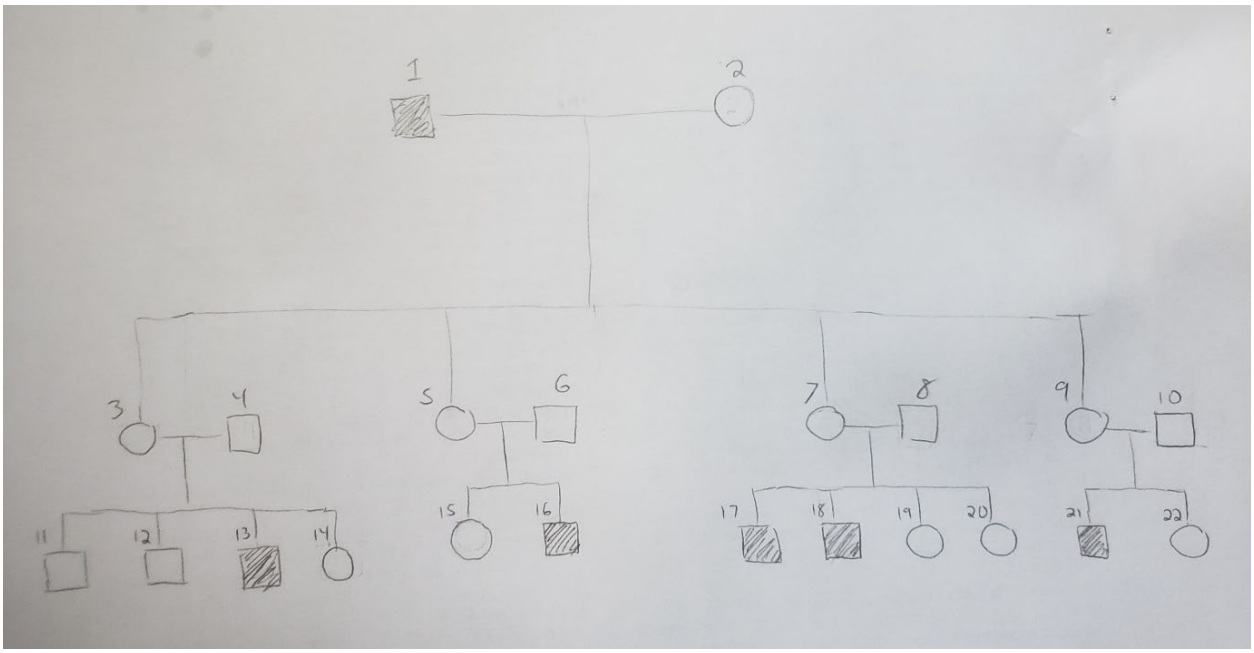
Use the hw1_family.csv file. The coding for the disease status is

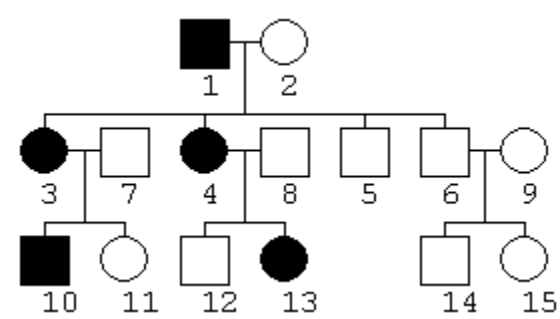
0 = unknown
1 = unaffected
2 = affected

Draw by hand a pedigree drawing of family 1, making use of appropriate standard nomenclature and shading individuals who are affected. What is the most likely form of inheritance of the trait (autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, complex inheritance). Assume a rare disease. Give an explanation of your choice.



This disease is mostly likely X-linked recessive. Every affected individual is a male. Individual #1 passed on the mutated X allele to all his daughters, so they became carriers. All the grandsons who were affected received their diseased alleles from their carrier mothers.

Using the pedigree below, assume the disease is inherited as an autosomal recessive with full penetrance. Full penetrance means individuals with the disease-causing mutation will always be affected.



Part A

Identify unaffected individuals who must carry one copy of the disease allele (obligate carriers).

Individuals 2, 5, 6, 7, 8 11, and 12 are obligate carriers.

Since the disease is autosomal recessive, Individual 1's genotype must be *dd*. All his affected children must have gotten one diseased allele from each and his mother. Therefore, Individual 2 must be a carrier. Both of the father's alleles are mutated, so his unaffected children (Individuals 5 and 6) must have one of his mutated alleles too.

The marry-ins (Individuals 7 and 8) must also be carriers because they each have one children who are heterozygous for the recessive diseased allele. Individuals 11 and 12 must also have at least one of their affected mothers' mutated alleles, so are obligate carriers.

Part B

Assume the disease is inherited as an autosomal recessive with full penetrance. What is the probability that individual 14 carries one copy of the disease allele? Assume that the disease allele is rare, and that ID 9 is not a carrier.

Individual 6 must be heterozygous for the disease. A Punnett square between a heterozygous and homozygous dominant results in a 50% chance of being homozygous dominant and 50% chance of being a carrier for the disease. There is a 50% chance that Individual 14 carries one copy of the diseased allele.

Part C

Assume the disease is X linked dominant with full penetrance. What is the probability that individual 6 carries one copy of the disease allele?

If the disease is X-linked dominant, all males with the diseased allele they inherited from their mother will be affected because they have no other X-chromosome to cover up the phenotype. Individual 6 is not affected, so there a 0% chance that Individual 6 carries one copy of the diseased allele.

Use the hw1_data.csv. The coding for the disease status is

0 = unknown
1 = unaffected
2 = affected

Part A

Fill in the following table.

Number of individuals	234
Number of families	50
Number of individuals with non-missing phenotype	209
Proportion of individuals with unknown phenotype	10.68%
Number of individuals with genotypes available	206
Proportion of individuals with any genotypes available	88.03%
Proportion of parents (founders) with missing genotypes	27%
Proportion of mothers with missing genotypes	36%
Proportion of children with missing genotypes	0.75%
Number of individuals with both known phenotype and non-missing genotypes	204
Average sibship size	2.68
Average sibship size (only children with both genotypes and known phenotype)	2.66
Number of families with at least 1 affected child	50
Number of families with at least 2 affected children	8
Proportion of affected parents (only parents with known phenotype)	29.33%

Part B

Provide an explanation for the large difference in proportion with missing genotypes between parents and children.

The founders/linker individuals/parents must be included in the dataset in order to link up the siblings and other relationship about the family in question. However, those founders are often not part of the study, so no phenotype or genotype information is available for them. That is why the proportion of parents with missing genotypes is much larger than the proportion of children with missing genotypes (27% of parents compared to 0.75% of children.)

Part C

Present a table with the genotype counts and proportions (frequencies) of all possible rs12075 genotypes. What is the most frequent genotype? Least Frequent? What is the symbol used for missing data?

A/A	A/G	G/G
118	78	10
57.28%	37.86%	4.85%

The most frequent genotype is A/A and the least frequent is G/G.
"./." is the symbol for missing data, which was the case for 28 people.

Part D

Present a table with the allele counts and proportions (frequencies) for SNP rs12075.

A	G
314	98
76.21%	23.79%

56 individuals had missing alleles.