Estimation of Parentage and LOD scores

Background from lecture notes

Consider an ordered triplet of genotypes (g_B, g_C, g_D) at a single locus for the 3 sturgeon (B, C, D). We are interested in identifying triplets consisting of an offspring and the maternal and paternal parent. The statistical properties of triplets under different relational situations are:

1. relationship UU: B, C, and D are unrelated (or not closely related) and thus the triplet contains neither parent of B
2. Relationship QU: D is the parent of B, but C is unrelated and thus the triplet contains one parent of B.
3. relationship QQ: C and D are both the parents of B and thus the triplet contains both parents of B.

The probabilities of these triplets will be denoted $P(g_B,g_C,g_D|R)$ where R=UU, QU, or QQ denotes the true relationship between individuals B, C and D.

$P(g_B,g_C,g_D|UU) = P(g_B)*P(g_C)*P(g_D)$

$P(g_B,g_C,g_D|QU) = P(\text{offspring } g_B|\text{parent } g_D)*P(g_C)*P(g_D)$
$= T(g_B|g_D,\text{--})*P(g_C)*P(g_D)$

$P(g_B,g_C,g_D|QQ) = P(\text{offspring } g_B|\text{parents } g_C,g_D)*P(g_C)*P(g_D)$
$= T(g_B|g_C,g_D)*P(g_C)*P(g_D)$

where $P(g_i)$ is the expected frequency of the i^{th} genotype under Hardy Weinberg in a randomly mating population and T denotes transition probabilities from parents to offspring.

There are 2 examples

Use the algebra example in the accompanying power point presentation as a guide to solve the problem.

In example 1, consider a locus with three equally frequency alleles segregating in the population [p1=p2=p3=0.33 for alleles a1, a2, and a3, respectively]

Progeny genotype (g_B=a1a3)

Mother genotype (g_C=a1a1)

Possible fathers genotypes g_D=a2a3 (dad 1), a3a3 (dad 2)

In example 2, consider a locus with 4 equally frequency alleles in the population [p1=p2=p4=p5=0.24 for alleles b1, b2, b4, and b5 respectively and the frequency of the b3 allele is much lower (p3=0.04)]

Progeny genotype (g_B=b1b3)

Mother genotype (g_C=b1b1)

Possible fathers genotypes g_D=b2b3 (dad 1), b3b3 (dad 2)

1. Answer questions a, b, and c for each example
    a) for each male genotype (a2a3 and a3a3) estimate the probability of observing the triplet genotypes under relationship UU and QQ.
    b) calculate the LOD score for each male as the natural log of the ratio (prob QQ/prob UU).
    c) which male is the most likely father?
2. Contrast the LOD scores between example 1 and 2. Why are the probabilities of each relationship and the LOD scores different when markers with different number of alleles are used?