Identity by descent

## Introduction

Self-fertilization is, of course, only one example of the general phenomenon of inbreeding—non-random mating in which individuals mate with close relatives more often than expected at random. We've already seen that the consequences of inbreeding can be described in terms of the inbreeding coefficient, f and I've introduced you to two ways in which f can be defined.[1] I'm about to introduce you to one more.

> Two alleles at a single locus are *identical by descent* if the are identical copies of the same allele in some earlier generation, i.e., both are copies that arose by DNA replication from the same ancestral sequence without any intervening mutation.
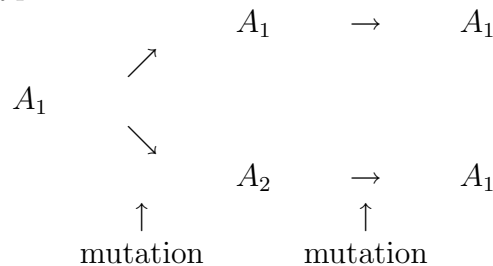
We're more used to classifying alleles by type than by descent. All though we don't usually say it explicitly we regard two alleles as the "same," i.e., identical by type, if they have the same phenotypic effects. Whether or not two alleles are identical by descent, however, is a property of their genealogical history. Consider the following two scenarios:

Identity by descent

$$A_1 \quad \rightarrow \quad A_1$$
$$\nearrow$$
$$A_1$$
$$\searrow$$
$$A_1 \quad \rightarrow \quad A_1$$

---
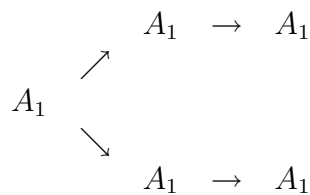
[1] In lecture we talked about f as a measure of how far genotype frequencies depart from Hardy-Weinberg expectations. In the notes for 30 January I point out that f can also refer to the *equilibrium* departure from Hardy-Weinberg expectations attained as a result of a particular pattern of inbreeding.

Identity by type

$$
\begin{array}{ccc}
 & A_1 & \to & A_1 \\
A_1 \quad \nearrow & & & \\
\quad \searrow & A_2 & \to & A_1 \\
& \uparrow & & \uparrow \\
& \text{mutation} & & \text{mutation}
\end{array}
$$

In both scenarios, the alleles at the end of the process are identical in type, i.e., they're both A_1 alleles. In the second scenario, however, they are identical in type only because one of the alleles has two mutations in its history.[2] So alleles that are identical by descent will also be identical by type, but alleles that are identical by type need not be identical by descent.[3]

A third definition for f is the probability that two alleles *chosen at random* are identical by descent. Of course, there are several aspects to this definition that need to be spelled out more explicitly.

- In what sense are the alleles chosen at random, within an individual, within a particular population, within a particular set of populations?

- How far back do we trace the ancestry of alleles to determine whether they're identical by descent? Two alleles that are identical by type may not share a common ancestor if we trace their ancestry only 20 generations, but they may share a common ancestor if we trace their ancestry back 1000 generations and neither may have undergone any mutations since they diverged from one another.

Let's imagine for a moment, however, that we've traced back the ancestry of all alleles in a particular population far enough to be able to say that if they're identical by type they're also identical by descent. Then we can write down the genotype frequencies in this population once we know f, where we define f as the probability that two alleles chosen at random in this population are identical by descent:

$$
\begin{align}
x_{11} &= p^2(1-f) + fp \tag{1} \\
x_{12} &= 2pq(1-f) \tag{2} \\
x_{22} &= q^2(1-f) + fq \quad . \tag{3}
\end{align}
$$

---

[2]Notice that we could have had each allele mutate independently to A_2.

[3]Systematists in the audience will recognize this as the problem of homoplasy.

It may not be immediately apparent, but you've actually seen these equations before in a different form. Since $p - p^2 = p(1-p) = pq$ and $q - q^2 = q(1-q) = pq$ these equations can be rewritten as

$$
\begin{align}
x_{11} &= p^2 + fpq \tag{4} \\
x_{12} &= 2pq(1-f) \tag{5} \\
x_{22} &= q^2 + fpq \quad . \tag{6}
\end{align}
$$

You can probably see why population geneticists tend to play fast and loose with the definitions. *If* we ignore the distinction between identity by type and identity by descent, then the equations we used earlier to show the relationship between genotype frequencies, allele frequencies, and f (defined as a measure of departure from Hardy-Weinberg expectations) are identical to those used to show the relationship between genotype frequencies, allele frequencies, and f (defined as a the probability that two randomly chosen alleles in the population are identical by descent).

Creative Commons License