The Confusing Vocabulary of Population Genetics

Alan R. Rogers

September 6, 2024

The modern literature of population genetics is confusing because two systems of vocabulary are in use, and these involve inconsistent definitions of the terms *gene* and *allele*. To avoid confusion, Jon Seger and I have used a third system in our course on Human Evolutionary Genetics. These systems are summarized in Table 1. System 1 is the one used in the classical literature. System 2 began to appear in the 1980s and is common today. System 3 is the one Jon and I use in our class.

As an example of the classical vocabulary in use, I offer Thomas Hunt Morgan [3, p. 25], who wrote that

> the characters of the individual are referrable to paired elements (genes) in the germinal material.

For Morgan, the genotype at each locus consisted of a *pair* of genes. He thought of genes as physical objects, which are transmitted from parent to offspring. R.A. Fisher [1, p. 8] made this even clearer a few years later.

> Those organisms (homozygotes) which received like genes, in any pair of corresponding loci[1], from their two parents, would necessarily hand on genes of this kind to all of their offspring alike; whereas those (heterozygotes) which received from their two parents genes of different kinds...

In the 1960s, people began using the term "gene" in a new way. Before that, a "gene" was the unknown hereditary something responsible for some difference between individuals—for example, round versus wrinkled in Mendel's peas. But in the 1960s and 1970s, "gene" came to mean something more specific: the DNA that encodes a particular protein [4].

This was unfortunate for at least two reasons. First, we now know that there are "genes" in the classical sense that do not encode protein. For example, some are "binding sites," which allow an enzyme to attach to the DNA. Others encode RNA that is never translated into protein. These are genes in the classical sense because a mutation in one can have phenotypic consequences. But they do not encode protein and are therefore not genes in the modern sense.

[1] The plural of "locus" is "loci," pronounced "low sigh."

Table 1: Three systems of vocabulary

	1	2	3
Position on chromosome	locus	locus	locus
Protein-coding locus	gene	gene	gene
Physical copy of DNA at locus	gene	allele	gene copy
One of several variants at a locus	allele	allele	allele

Second, people began to be uncomfortable with sentences that would previously have been unremarkable. For example, "at each locus we inherit one gene from our mother and another from our father." To geneticists trained in molecular biology but not in population genetics, this must have seemed wrong. At a given locus, the DNA we get from Mom and Dad represent *two copies of a single gene*, because both copies encode the same protein. It therefore seemed wrong to talk about getting two genes, one from Mom and one from Dad. This left us with no word for the physical thing that one inherits from Mom or from Dad at a particular locus. To fill this gap, the word "allele" was pressed into service. In this new system of vocabulary, "allele" means two different things, as shown in Table 1.

Here is how John Gillespie [2, p. 6] summarized the second system of vocabulary:

> Here we will use *locus* to refer to the place on a chromosome where an *allele* resides. An *allele* is just a bit of DNA at that place. A locus is a template for an allele. An allele is an instantiation of a locus. A locus is not a tangible thing; rather, it is a map describing where to find a tangible thing, an allele, on a chromosome. (Some books use *gene* as a synonym for our *allele*. However, *gene* has been used in so many different contexts that it is not very useful for our purposes.) With this convention, a diploid individual may be said to have two alleles at a particular autosomal locus, one from its mother and the other from its father.

Vocabulary systems 1 and 2 both use a single word to mean two different things. In system 1, that word is "gene;" in system 2, it is "allele." Unfortunately, the ambiguity inherent in system 2 is confusing more often than is that in system 1. For example, here is the same sentence in each of the three systems:

1. If the *genes* you inherit from Mom and Dad are different alleles, then you are a heterozygote.
2. If the *alleles* you inherit from Mom and Dad are different alleles, then you are a heterozygote.
3. If the *gene copies* you inherit from Mom and Dad are different alleles, then you are a heterozygote.

Our own vocabulary (system 3) isn't perfect either. We often find ourselves referring to a "copy of a gene copy" for example. All in all, I prefer system 1 to either alternative.

References

[1] Ronald A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.

[2] John H. Gillespie. *Population Genetics: A Concise Guide*. 2nd. Baltimore: Johns Hopkins University Press, 2004.

[3] Thomas Hunt Morgan. *The Theory of the Gene*. New Haven: Yale University Press, 1926.

[4] James D. Watson. *The Molecular Biology of the Gene*. New York: W.A. Benjamin, 1965.