APPENDIX 1 – SIMPLE NOMENCLATURE

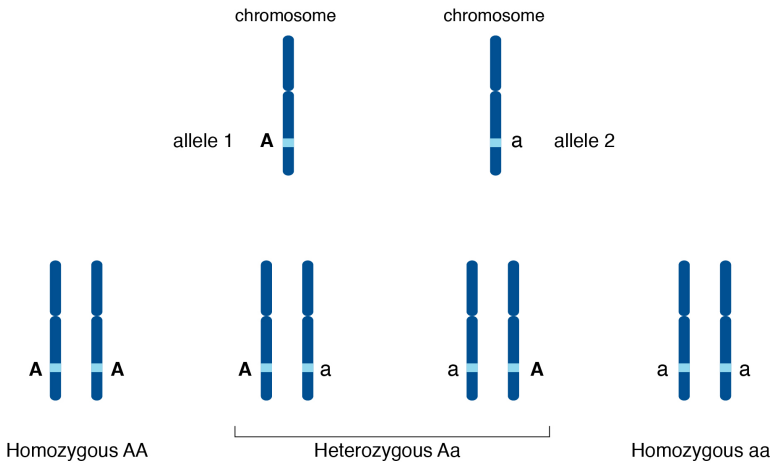


Figure 1.

Two chromosomes with the same gene but different alleles, *A* and *a*.

(Image used with permission of www.genome.gov provided it includes the link http://xoax.net/mediaGallery/Allele-dat-342)

INTRODUCTION

You have probably encountered symbols in high school genetics and perhaps in previous chapters assigned to you in this course. Chapter 1 points out that a position on a chromosome is a gene locus, which simply means the location at which a gene is found. At that position, the gene could be one of several variants that we call alleles. The set of alleles comprise an organism's genotype, which might affect the organism's phenotype through the action of the genotype.

When you are solving genetics problems you have to keep track of different alleles of one or more genes. This can become confusing, so understanding how a gene works, how alleles vary, how they are expressed, and the influence they have on phenotype is important. A consistent nomenclature system – a set of rules dictating how you name alleles – becomes immensely helpful. Online Open Genetics has an activity that demonstrates why a good nomenclature system is helpful. The icon to the right indicates an online module.

1.1 Power of Labels

1. WHY USE NOMENCLATURE?

Word and practical problems in biology can get confusing in a hurry, particularly if you're distracted by something like exam stress! When you are manipulating several ideas, it is good practice to be thoughtful and follow rules that keep you consistent in your interpretation. As an instructor, I have seen work in which the student clearly got flustered and forgot that the mutation he or she was working on was dominant. This often leads to an answer that is inconsistent with the data.

There are a few simple rules we can use for nomenclature. This first appendix uses a simplified system to communicate allele characteristics. The next appendix will show a more complicated system that carries even more information in the gene symbols.

2. THE BIOCHEMISTRY OF GENE BEHAVIOUR

2.1. THE MECHANICS OF GENE EXPRESSION

We often think of genes as "made of DNA"; they reside in the nucleus and endosymbiotic organelles of eukaryotes or the nucleoid region of prokaryotes. As described in Chapter 13, they are transcribed into an RNA message by RNA polymerase then interpreted by ribosomes that assemble particular amino acids into a polypeptide strand (also known as a protein) based on the sequence of nucleotides. In a cell, proteins can act as enzymes, structural features, pigments, and a host of other functions, including regulating the expression of other genes. This expression of genes leads to how an organism looks – its phenotype.

2.2. DOMINANT AND RECESSIVE; HOMOZYGOUS, HETEROZYGOUS, AND HEMIZYGOUS

Alleles themselves do not directly exert an effect on the phenotype of an organism. You will recall from Chapter 3 that genes are instructions, often for proteins. It is the effect of the protein that causes an organism to take on particular traits. In this chapter we will look at how to design symbols appropriate to communicate the characteristics of alleles you want to investigate, but keep in mind that the allele instructs what kind of protein to make.

Chapter 1 points out that organisms usually fall into the classes of being diploid or haploid. Humans and eukaryotic genetic systems usually assume the organism is diploid, which means that most chromosomes are represented as pairs. Each pair has homologous loci: the term *homologous* means "the same information" and refers to the gene at each locus. Note that although the loci, or genes, are the same, the alleles that comprise them may be different!

If an organism is **diploid** and **homozygous** for an allele (Figure 2, left and right), the gene at the same position of the homologous chromosomes is the same allele. Only one type of protein is made. If an organism is diploid and **heterozygous** (Figure 2, middle), and the protein from one allele influences the phenotype more than the protein from the other allele, we use the terms **dominant** and **recessive**, respectively.

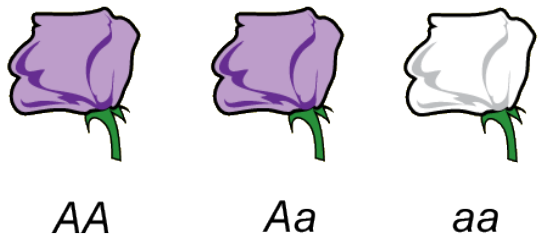


Figure 2.

Relationship between genotype and phenotype for an allele that is completely dominant to another allele.

(Original-M. Deyholos -CC:AN – from chapter 13)

If one copy of an allele makes enough protein to compensate for the absence of protein from the other allele in a heterozygote, it will influence the phenotype. If this phenotype looks identical to that of an organism homozygous for the "functional" allele, we consider the "functional" allele to be *dominant* to the "null". We could also say the "null" is recessive to the "functional" allele. Keep in mind that the alleles themselves aren't doing anything, but it's common practice to label the alleles as dominant or recessive, although in reality we are talking about the expression of those alleles.

1.2 Biochemistry of Gene Behaviour

Why this is important is how the proteins from two alleles interact. If both proteins are identical (from a homozygous genotype) the phenotype that results will be that of the action of one "type" of protein, even if though there are two copies of the gene – they are both the same allele. There's no real interaction because the proteins do the same thing. It's in heterozygotes that we can see whether a particular allele is dominant over another or otherwise influences the phenotype in an interesting way.

Here's another point about the "normal" allele: it was honed by natural selection over a long period of time. Genes are instructions for the protein tools of an organism's cells. For this reason we often call the "normal" allele of a gene the "wild type" allele. This would be the allele most common

in the wild, presumably because it provides a benefit to the organism. Thus, most mutations are likely to reduce the effectiveness of the wild type allele, although the process of evolution allows (and, in fact, requires) an occasional beneficial allele to permeate a population if it provides a selective advantage.

Note that we don't use the terms homozygous or heterozygous in *haploid* organisms. If they are haploid, their phenotype will reflect the genotype of the only allele present. The proper term for their genotype is *hemizygous* to reflect only one copy of each gene.

2.3. THERE ARE MANY KINDS OF ALLELES FOR A GENE

A mutation is a change in nucleotide sequence. Chapter 11 goes into more detail of what this means. What's important now is that you understand that the amino acids of a protein can be different if we compare different alleles of a gene and they may behave differently – often one protein will "work better" than the other. If the promoter of a "functional" allele of a gene is damaged, the allele that is created might not even create an mRNA so no protein will be encoded by that allele. This is called an amorphic or "null" mutant (See Chapter 1).

Keep also in mind that a gene can be mutated in different parts of its sequence to create different alleles. A diploid organism can have a maximum of two alleles (aside from gene duplication or abnormal chromosome structure, but ignore that for now). In a population, though, there can be many, many different alleles. Perhaps allele one decreases protein function and allele 2 is even less effective. This describes an **allelic series** as follows: wild-type > allele 1 > allele 2 > any null allele. Wild-type alleles in this case encode the most effective protein. Null alleles represent catastrophic mutations that eliminate transcription or produces proteins that can't function at all. See section 3.2 on page 4 below for another example of an allelic series.

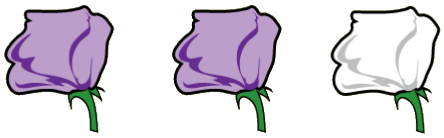
2.4. GENDER AND GENE INTERACTIONS

Finally, here are a couple points about other interactions before we move on to gene symbols. Some organisms have a pair of sex chromosomes that dictate gender (see **Chapter 4**). We'll ignore incorporating ways to communicate if a chromosome is a sex chromosome in this chapter; **Appendix 2** will show how we can indicate those using a symbol.

3. BASIC NOMENCLATURE

3.1. SINGLE LETTER SYSTEM

Sometimes what you want to do is a little rough work for investigating your genetic model. A genetic model is a diagram of the logic that you propose for a particular type of inheritance. For instance, if you cross a true-breeding purple plant with a true-breeding white plant (*e.g.* see Figure 2 – again – to the left. It was first shown on page 2; cross the outer two plants) you will get a heterozygote (the middle plant). If we name the gene after the mutation (*a* is the first letter in "albino"), we know that the heterozygote will have one capital letter "*A*" and a lower case "*a*". The heterozygote is the F_1 generation ("first filial", which means it's the first child from parents that are crossed). The F_1 is purple, which means the "*a*" allele is recessive; only one copy of the "*A*" allele is needed for enough purple pigment to make it identical to one true-breeding parent. This is complete dominance.



We can't know from the information given which allele is wild type or mutant. One hypothesis, though, is that purple pigments are required to attract pollinators and therefore would help the plant in the wild. Albino plants could be a mutant and might not generate as many seeds for lack of pollinators. If the context of your genetics problem doesn't indicate which allele is wild-type, it's good practice to name your allele based on the recessive trait. Often the recessive allele is the mutation. We might consider that the "*a*" allele is null and makes no pigment at all. One or two "*A*" alleles make enough protein to cause the plant to be purple.

Figure 2 already assumes that the capital letter (*A*) stands for an allele that encodes a protein for purple pigment and the recessive allele (*a*) doesn't make pigment. Thus the *Aa* heterozygote is sufficient evidence to adopt upper- and lower-case letter "*A*"s to communicate the characteristics of purple and white alleles.

A note of caution. When you're writing down gene symbols for homework or on an exam, be sure to make the characters distinct. A typewritten "*y*" is easy to distinguish from the upper case "*Y*" but not as easy when writing it down. Instructors who ask you to show your work need to be able to follow your logic train. More important than that is YOU have to be able to follow your own reasoning. Students often switch symbols and come up with an answer that is inconsistent with the data given because of this. Consider underlining your capitals or putting a line through one of them to make it distinct (*e.g.* ~~*Y*~~ for the dominant allele; *y* for the recessive).

Name the gene after the mutant phenotype

Some instructors would accept "*P*" for "*purple*" for the previous cross. However, the better answer is to follow an established system. During "exam fog", it's easy to get lost if you are inconsistent with how you develop your symbols. During your study period and when you're practicing genetics problems, be thoughtful about the gene names you choose.

Let's always choose a letter based on the mutant phenotype for our gene symbol. If we are presented with a ladybug mutant that is small, we might choose "*d*" for "*dwarf*". Geneticists sometimes set up a research program based on unusual phenotypes of the organism they are studying. The fact that a mutant phenotype that is heritable exists tells us that there **is** a genetic control for the trait and that it might be isolated in the lab. When you look at your classmates, you don't necessarily note that none of them has an arm growing from the tops of their heads. If one student had this trait, however, you couldn't help but notice it. If you found out it showed up in that student's ancestry in a predictable fashion, you might reasonably suggest that there is a genetic basis for that. If it happened to be controlled by a single gene, you might call the gene "*extra arm*" or "*arm head*". If it happened to be a dominant trait, you might use the letter "~~*X*~~" or "*A*" for the mutant allele. The wild type allele would be "*x*" or "*a*", respectively.

If you don't know which is mutant, use the recessive trait for the gene name

What if you don't know which allele is mutant? What if you're presented with two true-breeding frogs: one that is gold and one that is yellow. If you don't know what the predominant colour in nature is you can't know which one is mutant. If you crossed them and all the progeny are gold, then you know the dominant allele encodes a protein to make it *gold*. The recessive, therefore, is "*yellow*" and you should name the gene "*y*" after the recessive phenotype. This means the dominant allele would be "*Y*". Your offspring would therefore be *Yy* and the gold parent would be *YY*. The yellow parent would be *yy*.

1.3 One Letter System

Apply these ideas at the Online Open Genetics exercises.

3.2. SUPERSCRIPTS

Sometimes a letter is used as the name of a gene, and superscripts can modify it to indicate the different alleles. One common single letter code for an allelic series is "*I*". Red blood cells can have their cell membranes modified by sugar tags that give rise to our blood type. One allele of *I* gives rise to blood type A and is therefore called I^A. An enzyme encoded by I^B modifies sugars to create blood type B. A heterozygote I^AI^B demonstrates *both* sugar tags because those alleles are expressed – they are codominant. People with blood type O only possess alleles for the *I* gene that don't work and are therefore recessive – they don't modify the extracellular sugar tags. Because it is recessive, individuals are homozygous for *i*: they are *ii*. Chapter 13, Section 7 has more detail on this allelic series.

Sometimes a superscript "plus sign" is used to denote the wild type allele. One might use the symbol W^+ to indicate a wild-type allele that promotes wing growth. Note that the generic "wing" gene name isn't a best practice – name the gene after the mutant phenotype! A wingless mutant would be W^-. You should never use a "+" and shift the case of the letter unless you are dealing with a special case such as the codominance in the blood type example above. The capital "*I*" letter indicates it is dominant to "*i*". The superscript A and B for the codominant alleles indicate the dominant alleles are different from each other.

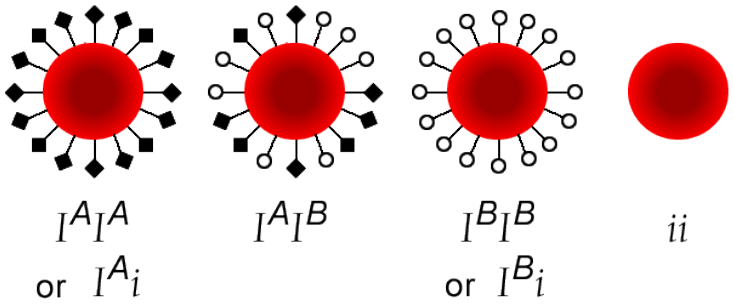


Figure 3.

Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the *i* allele. The I^B allele is completely dominant to the *i* allele.
(Original-Deholos -CC:AN – from Chapter 1)

Superscripts can be symbols, a single letter, or many letters. They modify the gene name only in the superscripted symbols: the regular-sized letters are identical between them (see Table 1). This means that Abc^+ and *abc* would be different genes (*i.e.* not allelic); Abc^+ and *Abc* are alleles, as are abc^+ and *abc*. Note that a superscript is not mandatory for all alleles of that gene, depending on the convention.

Alleles of bacterial genes are typically indicated with a superscript + or -. For example, a bacterial allele that creates an enzyme that makes methionine would be met^+, and a defective allele of that gene is met^-.

Table 1. Examples of genes using a superscript modifier.

$white^{apr}$ or $white^{apricot}$	An allele of the *white* gene which has an "apricot" phenotype
Abc^+ *Abc*	Two alleles for the *Abc* gene (wild-type and mutant, respectively). Note the mutant allele is dominant.
w^+; w^a; *w*	Three alleles in a series for the *w* gene. The first is wild type; the second two are different mutant alleles.
bio^+; bio^-	A wild-type allele of a *biotin* gene and its recessive counterpart. This is likely a bacterial gene due to the convention.

4. THREE-LETTER SYMBOLS

It's perfectly acceptable to use a single letter or even two letters. Sometimes, though, multiple traits spelled with the same first letter can get confusing. Using three letters for a gene symbol can make it easier to remember what the letters stand for. In fact, for some model systems those who study them adopt a defined nomenclature system. A plant often used for genetic studies, called *Arabidopsis thaliana*, has a three-letter code (`https://www.arabidopsis.org/portals/nomenclature/namerule.jsp`).

Just as we saw for the one-letter symbols, the dominant allele has the first letter capitalized and the last two letters are lower-case. Recessive alleles are all lower-case. With three letters, you can make gene names that are easier to keep track of. For example, you might see a fly with an extra set of wings. Instead of calling it "*w*" for "wings" (which is a poor choice because it doesn't represent the mutant phenotype), you can instead call it "*exw*" for "*ex*tra *w*ings". Then, when you see it, you can sound out the abbreviation and remember that it stands for the mutation. If the wild-type allele for this is dominant, then you would write that one "*Exw*". The first letter indicates that it is the dominant allele. What you can't tell just from these examples is that wild-type allele is dominant! Appendix 2 will go into an extension of this system so

you can add that information to the gene symbol. *Hint:* remember in Section 3.2 above that the "+" superscript indicates the wild-type allele.

5. LINKED GENES

Mendel was lucky. He studied a variety of traits in pea plants and his data were consistent with his idea of traits being encoded by pairs of discrete heritable units. He didn't call these "genes" and had no idea about their chromosomal origins or chemical makeup. It turns out that the genes he studied were either on different chromosomes and so assorted independently (See Chapter 9), or so far apart on the same chromosome that linkage could not be detected (See Chapter 10).

Symbols for a gene can be drawn on a page to communicate their position on a chromosome. To do this, we use a forward slash (/) to demonstrate what is on each chromosome. Figure 4 shows how we might conceptualize the position of a gene on two chromosomes by collapsing the chromosomes into a single line.

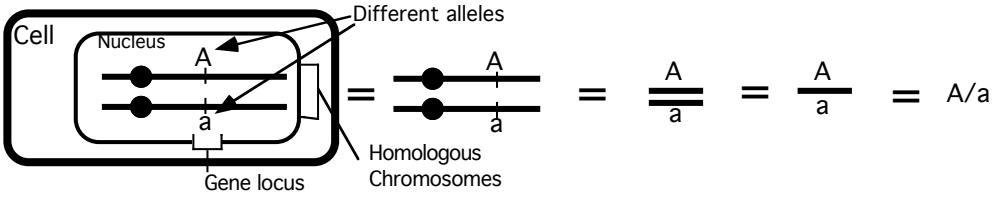


Figure 4.
A diagram of how chromosomes, loci and alleles look in the cell, and how we depict them in written form. Note that only one gene is represented in this figure and chromosomes have not replicated.

(Original-J.Locke- CC BY-NC 3.0 from Chapter 10)

There's no question about where the gene is located when only one trait is under investigation: it will be at the same position on each homolog. Two genes, however, can be one of three possibilities. Each possibility has implications for gene mapping and predicting ratios from a dihybrid cross. Figure 5 shows the positions of genes for an unlinked situation as well as linked genes in coupling and repulsion configurations. If genes are unlinked, put the allele symbols for one gene on either side of one slash followed by a semicolon (indicating that it's unlinked) and the other gene with the alleles separated by a second slash (*A/a; B/b*). When genes are linked, only one slash is used: remember, the slash stands for a pair of homologous chromosomes. Genes in coupling would have the dominant genes together on one side of the slash and recessives on the other side (*AB/ab*). Repulsion would represent the other arrangement (*Ab/aB*).

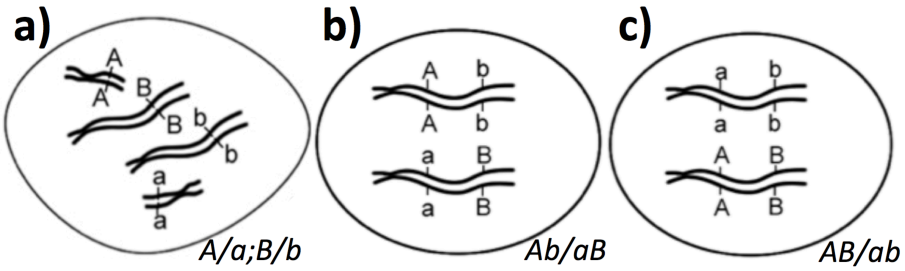


Figure 5.
Three gene arrangements for cells of genotype AaBb. Chromosomes are replicated (shown with sister chromatids). a) demonstrates unlinked genes: *A/a; B/b*. Linked genes are shown in b) repulsion *Ab/aB* and c) coupling *AB/ab*.

(Modified by T. Nickle from an original by J.Locke- CC BY-NC 3.0 from Chapter 18)

Practice your skills with identifying linked and unlinked genes online in module 1.5. Some examples of different forms of gene symbols are shown in Table 2. Keep in mind that sometimes you have flexibility in which system of nomenclature you use, but sometimes it is dictated to you, for example in publications or other formal submissions. You are discouraged from inventing your own system or mixing up different systems because it will confuse your readers (or graders!).

1.5 Symbols for Linked and Unlinked Genes

A more advanced system of nomenclature is outlined in Appendix 2. New rules are introduced to help you identify sex linked genes and predict phenotypes from gene symbols alone.

SUMMARY:

- Symbols are used to denote the alleles, or genotype, of a locus.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- In a diploid organism, the alleles can be homozygous, heterozygous or hemizygous.
- Allelic interactions at a locus can be described as dominant vs. recessive, incomplete dominance, or co-dominance.
- Two different genes can be on the same chromosome (linked) or on different ones (unlinked)
- Steps:
    - Can you identify the mutant? Name the gene after its phenotype
    - If you cannot tell which allele is mutant, name the gene after the recessive allele
    - Capitalize the first letter for the dominant allele; use lower-case for the recessive allele
    - If genes are linked, write the gene symbols together on each side of the slash. If genes are unlinked, they appear on different sides of a semicolon

Table 2. Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used in publications because not all alleles show complete dominance and many genes have more than two alleles. It's quick and easy for you to use when working out genetics problems when you are sure each gene involves only two alleles.
a^+ and *a*	Superscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +. The mutant allele of gene *a* would be recessive.
met^+ *and* met^-	This is typical of a prokaryote gene symbol. It could be referring to wild-type (functional) and mutant (nonfunctional) alleles of a gene that makes a protein in the methionine synthesis pathway.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes. Both representations in this row are identical: it represents a homozygous dominant.
Aa/Aa or *Aa/aa*	Note that this example shows two alleles of the gene *Aa*. We know that the gene symbol is two letters because the slash separates the allele found on each of the homologous chromosomes. We cannot tell if the mutant phenotype is recessive because there's no indication which is wild type.
Grn^+ *shr/Grn shr*	The three-letter system is used here. "*Grn*" might mean that the phenotype is "green", but we can't be sure. What we do know is that the mutant allele codes for a protein leading to a dominant phenotype. The wild-type allele must be recessive to the mutant allele. Maybe "*shr*" means "shrunken" or "short", but we know that the mutant phenotype can only be seen in the homozygous recessive configuration. The phenotype for this organism is mutant for both *Grn* and *shr* traits. Final note: the genes are on the same chromosome based on the position of the slash.
bob^+*/bob; mia/mia*	This also uses the three-letter system. The organism is heterozygous for *bob* but shows the wild-type trait in its phenotype. It is homozygous recessive for *mia* and therefore shows that mutant phenotype. The genes are unlinked.