1

Introduction

This book is intended to be a complete discussion of horse and donkey colors. It includes details of color identification as well as their genetic control. The goal is to include all color variations occurring throughout the world and to fit these into a framework that is based on traditional American nomenclature as well as on the genetic phenomena controlling the color variations.

Identification and definition of horse color are important for several reasons, and each of these reasons demands a different organization and presentation of the material. Reasons for accurate horse color classification include identification of individual horses for legal purposes, health records, and breed registrations. In addition, breeders who are interested in producing or avoiding specific colors of foals find that accurate identification of colors in their breeding stock is essential to their success.

The organizational structure of this book combines a strictly visual approach (what color the horse appears to be) with the genetic control (how that color was produced by the interactions of the genes involved). Each section starts off with the visual approach and then delves into the genetic aspects. Unfortunately, a few specific details of horse color are better understood by first explaining the underlying genetic mechanisms that give rise to the colors rather than from any other point of view. In those few cases genetic mechanisms are presented first with the visual aspects following.

The genetic approach to understanding horse color is becoming increasingly common as deoxyribonucleic acid (DNA)-based tests for many of the genes causing the colors are now available. The results of these tests help breeders to better understand the colors and the genetic basis of their production, as well as the range of colors a specific horse can produce. While genetic testing has been generally helpful, it has also revealed a few confusing issues. Some horse colors, when classed only by visual appearance, appear to be a single group, but this single group includes the results of several very different genetic formulas. This occurs with **black** horses, and also with some of the light colors such as **champagne** and **pearl**, as well as **silver dapple** and **mushroom**. These different genetic formulas are presented at length in the corresponding sections of the book and are examples of the complexity of the genetic systems that produce a horse's final color.

New instances of multiple genetic formulas leading to visually identical colors are regularly coming to the attention of researchers in this field, and these can easily cause confusion to owners of horses with some of the variants. These colors present a very real challenge because a single outward appearance can spring from very different genetic mechanisms. In most cases it is fortunate that only one of the several possible genetic mechanisms for a confusing set of colors is common, and the few others that are possible

Equine Color Genetics, Fourth Edition. D. Phillip Sponenberg and Rebecca Bellone.

are much more rare. Consequently, discussion of the colors can still proceed from the basics of the most common mechanisms, even though the more rare mechanisms that lead to similar visual results must also be considered for completeness. Those that are known are presented here.

Importantly, even though some mechanisms are common and others rare, when a rare mechanism is present it becomes the only one of importance in understanding how a specific horse or donkey will produce color in a breeding program. In a very real sense, "rare" is only really an issue at the population level. At the level of the individual horse the only thing that is important is the specific blend of individual genetic variants, regardless of their frequency in a population.

Choosing a set of nomenclature for horse colors presents an interesting array of challenges. Historical approaches were of necessity based solely on visual classification. These older, traditional systems were detailed and technical and have served well for centuries. They were based on a rich lore of horse-specific information and were generated by people closely familiar with horses and their variation. This traditional approach deserves great respect for having served well for so long.

From a strict and non-equine viewpoint, nearly all horse colors could simply be described as a shade of brown or black or a mixture of those two. People with no equine background do indeed tend to lump most horses as "**brown**," because to the inexperienced eye they indeed are. One step beyond the "most horses are **brown**" approach is the traditional equine-specific nomenclature based on various details related to visual appearance. While traditional nomenclature varies region to region, it has served well for identifying horses by color. It has also served as the framework under which genetic investigations were first accomplished.

In recent years, an approach based more on genetics has come into vogue for understanding and classifying horse color. This approach can often simplify nomenclature. This is especially true when similar phenotypes are caused by mutations in the same gene. However, the genetic approach can also complicate nomenclature, as novel genes or mutations are given new names that have no equivalent in the more traditional system. This can muddy the identification of horse colors.

The presence, for some colors, of multiple genetic mechanisms causing visually similar results presents very real problems for developing a consistent nomenclature based on genetic information alone. Nomenclature can either be visually based or genetically based, but choosing either one of these as the primary organizer of nomenclature will inevitably cause problems for situations in which the other basis is a more compelling consideration. Still, the historic, visually based approach is the only one that is likely to succeed in field situations where genetic details are only rarely known. In contrast, the genetically based approach can be more useful in classing and defining the color of breeding horses when their owners are preferentially interested in producing specific colors of foals. This guide generally uses a visually based approach, but the more genetically based approach is referred to when appropriate.

Colors are discussed in a sequence that first examines dark colors because these tend to be the most common colors in most breeds and most regions. An understanding of less common (generally lighter) colors is then built as a progression from the dark colors. Each section first defines and classifies a color or group of colors by the visual approach and then delves into the details of what is known about the genetic control.

The addition of white hairs can be superimposed on any background color, and each pattern of these white hairs or patches is examined after all basic colors are considered. The patterns of white hairs are organized in the same fashion of first classifying each pattern by its visual appearance, followed by an explanation of the underlying genetics.

Donkey colors are the subject of a separate discussion following the section on horse colors. The donkey color discussion is organized in a manner similar to that of the horse color section, but is shorter. Much less is known about donkey color than is known about horse color, although this body of knowledge has recently begun to expand. Subtle details of donkey color are understood more readily when considered in the light of horse color identification and genetics, which also makes a shorter discussion appropriate.

Mule colors are omitted, except for a few examples. This is due to mule colors being somewhat less well understood than those of horses and donkeys, while at the same time being generally consistent with the expected interactions of the genes controlling color in the two parent species.

A series of summary tables is presented in Chapter 13, after the text and illustrations. Table 13.1 is a list of color names that are included in the text, serving as an attempt at a reasonably complete single list of horse color names and their main distinguishing features. Table 13.2 is a similar list for the patterns of white hairs. Table 13.3 lists the various genes affecting horse color and their actions. This includes both the loci and alleles. Table 13.4 is similar to Table 13.3, but is devoted to the genes associated with patterns of white. Table 13.5 lists genotypes of the different colors so that breeders can more adequately understand them and predict the possible color outcomes from mating various colors of horses. Table 13.6 outlines the various alleles present in different breeds. It can be used by breeders to develop the potential array of colors in various breeds. Table 13.7 is a large and cumbersome table that outlines the potential results of mating various parental colors. Table 13.8 has the details for horse names, breeds, and sources of photographs and other figures.

1.1 Basic Horse Color Identification

One purpose for understanding horse color is to be able to identify horses accurately. Accurate identification of horse color is a key ingredient in understanding the genetic or biologic basis of color and is the foundation upon which genetic investigations are built. Even a casual observer soon realizes that horses have a wide variety of colors. A standardized classification is necessary to begin communicating subtle differences between some specific horse colors. Any standardized system of color nomenclature depends on observers viewing a horse's color in the same general way.

Different languages and cultures each have distinct approaches to describing and classifying colors of horses. These distinctions are due to differences in deciding which specific characteristics of color are most important. Different classification systems each proceed logically from a few key characteristics, although these characteristics vary from system to system. The approach of each language has merit, even though the internal logic differs from one to another. Languages and cultures tend to vary enough that a concise one-to-one correlation of color names is usually impossible between languages, as certain details that are important in some languages are simply lacking in others.

An ideal system of horse color nomenclature would be one in which each unique color name (phenotype) corresponds to a specific genotype (genetic makeup) and each specific genotype results in a unique color or phenotype. An absolutely perfect one-to-one correspondence between genotypes and phenotypic nomenclature is lacking in all systems that are in use. Sometimes the lack of correspondence is for biologic reasons, but more frequently it arises from cultural or historic reasons.

Even though all systems fail to accomplish a tight one-to-one correspondence of terminology and genetic foundation, it is important to acknowledge that all nomenclature systems have a cultural and historic backdrop and that each has merit for specific details. For example, the one color group designated as **chestnut** in English is seen as three different colors (**alazán**, **ruano**, and **tostado**) by some Spanish-speaking traditions.

Attempts to force a genetically based nomenclature onto descriptions of horse color are becoming increasingly common. Though these systems may have great utility for horse breeders interested in producing specific colors, it is also true that many horse owners and enthusiasts find them confusing. These newer systems have failed to be adopted for general use because older and time-tested systems of nomenclature based on visual appearance have been successfully used for millennia and are difficult to replace.

As already noted, some single colors result from distinct and different genetic mechanisms, and therefore a one-to-one correspondence of genotype and color is impossible without genetic testing. As an example, many combinations of different dilution mechanisms are notoriously consistent in producing beige horses with pale brown manes, tails, and lower legs, and yet each of these horses comes to its similar color through a different genetic combination. These pale horses, though all a similar color, will each produce a very different array of colors in their foals. Devising unique names for each unique genetic combination can be useful to breeders but ignores the fact that the horses being described are all remarkably similar colors when seen out in the field. This is especially true if the horses are viewed at a distance.

The similar visual results of multiple genetic mechanisms are doubly confusing if nomenclature separates them and demands documentation of genes and alleles for each horse before it can be classified by color. A strategy for compensating for the lack of a one-to-one correspondence between nomenclature and genetics is to note the multiple genotypes included under a color name wherever possible. Likewise, the reverse problem of a single genotype giving rise to colors that are assigned different color names can be noted. These confusing situations are fortunately rare, so that a general trend toward one-to-one correspondence of color name to genotype is indeed the case. Those cases in which multiple mechanisms exist usually consist of one very common mechanism and one or two additional, but very rare, ones.

A few concepts form the sound foundation from which horse color can be understood, regardless of the system of nomenclature in use. The first important concept to understand in horse color identification is that background colors of horses occur independently of dilutions, and are also independent of any white markings horses may have. White hairs occur as a result of hair lacking pigment granules, so white patches, markings, or individual white hairs result from absence of color rather than being a true color in themselves. An incorrect belief, held by many people, is that various colors are superimposed on a white horse in much the same way that an artist applies paint to a white canvas. Any white areas, they believe, simply did not receive color. This incorrect idea, therefore, holds that horses are basically white. However, the truth is just the

opposite; white is superimposed on and covers up areas genetically destined to be specific colors. The important detail is that the genetic actions determining the color of the pigmented areas generally operate independently of those determining the extent and location of the white areas.

It is very important that white be understood to be superimposed over some color that would otherwise have been present. All horses have the genetic capability to produce pigment over all the body. This capability has been changed on some horses (or on portions of some horses) by a superimposed genetic directive to impede the production of color, leaving white hairs or areas. The basic color of a horse must therefore be considered by first ignoring any white areas. Horses that are entirely white (or nearly so) will of course make this approach impossible. However, the tactic of first ignoring white does work well for most horses and is essential in deciding the basic color of a horse. For this reason, patterns of white are discussed in the chapters following those on basic colors because the two categories (color and patterns of white) are genetically distinct.

Another important concept is the definition of the "points" of a horse. In horse color terminology the points are the mane, tail, lower legs, and ear rims. The importance of the concept of the points is that their color usually determines the name given to the overall color combination on a horse. Specific combinations of point color and body color determine most horse color names.

The two main groups of horse colors are those with black points and those with nonblack points. Nonblack points are usually red or cream, but occasionally are a brown color. The division of points into black and nonblack is important for identification and also has important genetic implications. Specific combinations of point color with body color yield the final color name; so, once point color is appreciated, it becomes fairly easy to identify most horse colors.

Black and nonblack points are usually easy to distinguish from one another (Figures 1.1 and 1.2). In some instances black manes and tails become faded or sunburned to brown, and in these cases the lower leg is the most accurate indicator of point color. In most horses with black points the black carries to the hoof and involves at least the pastern. In most horses with nonblack points the pastern and coronary region are lighter than the remainder of the lower leg, so this region is very useful in deciding whether a horse has black or nonblack points. Distinguishing between the two groups of point color is usually simple because horses with confusing point color are rare.

Point color can be confusing on foals (Figure 1.3). Foals of all colors frequently have very pale points, even on those colors that have black points as adults. Even though experienced observers can usually predict adult coat color from characteristics of the foal coat, exceptions are numerous and frequent enough that everyone should be cautious in predicting adult color from foal coat color. Betting on the final color from a foal's coat is a good way to lose money, especially in breeds with wide color variation!

Also potentially confusing are horses with extensive white markings, because they can have their point color completely masked by the absence of pigment in these markings (Figure 1.4). In these cases mane and tail color become the most important indicator of point color, but even then it can still be difficult to accurately assess point color. It is essentially impossible to determine point color accurately on some horses with extensive white markings.

Various combinations of point colors and body colors are given different names in different geographic regions, and no single system or language is complete for naming

Figure 1.1 **Bay** horses have black points, which include the mane, tail, ear rims, and the lower portions of the legs. *Source:* courtesy of Dyan Westvang.

Figure 1.2 **Chestnut** horses have nonblack points. This horse has red points that are similar to the body color; however, point color can vary widely in horses with nonblack points. *Source:* courtesy of Dyan Westvang.

Figure 1.3 Foals of colors with black points, like this **dun**, usually do not have black points at birth but only develop them later. This is one reason adult colors can be difficult to predict from foal colors, even for experienced observers. *Source:* courtesy of Nancy Cerroni.

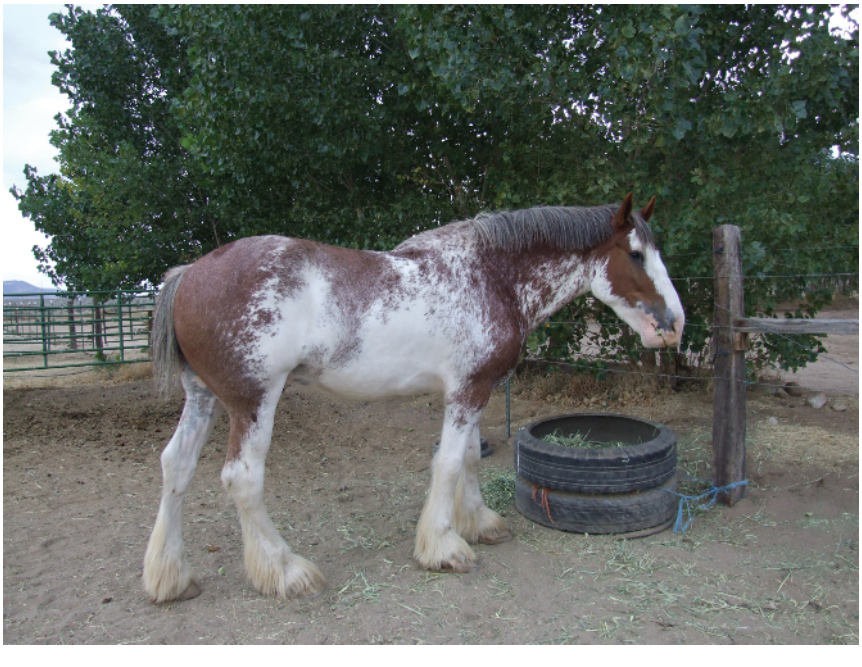

Figure 1.4 This **bay** horse has extensive white markings that obscure the lower leg color, leaving the mane and tail as the only accurate indicator of point color. On this horse, even the mane and tail are mixed with white hairs, although the basic black point color is still clearly evident. *Source:* courtesy of Jeannette Beranger.

the details of all of these combinations. The approach taken in this guide is consistent with the usual approach in the western USA. The westerners in the USA developed a fairly detailed vocabulary for describing horse color. This vocabulary functions well for nearly all color variations possible on horses. In addition, it usually corresponds well to the underlying genetic mechanisms. A few rare colors or combinations have no names in English. In such cases, other cultures or languages have been consulted for names and concepts that will help in understanding these rare colors and how they occur. While other languages and cultures have been considered, the final nomenclature is usually an English equivalent so that English speakers can more readily understand and use the concepts adopted from these sources.

A detail that can sometimes lead to confusion is that horses can vary in color from season to season or year to year (Figures 1.5 and 1.6). Horses are generally darkest when they shed their thick winter coats in the spring. Sun, wind, and rain can then act to bleach the color, although some horses remain unaffected by such weathering. Horses also can change shade of color due to their state of nutrition, condition, and general health. Healthy, well-fed, and well-conditioned horses are usually darker than are those less fortunate horses lacking the same benefits.

An additional detail that becomes a problem confronting an observer of horse color is the inescapable fact that every color varies over a range from light to dark. It is therefore always possible to find some individual horses that are at the boundary of two defined colors. This is simply a consequence of the complexity of the genetic control of horse color, coupled with the uncertainties and inconsistencies of the environment. A

Figure 1.5 This is a **black tobiano** horse early in spring after shedding the winter coat, when colors are usually at their darkest and most distinct. *Source:* courtesy of Dyan Westvang.

Figure 1.6 This is the same horse as in Figure 1.5, but in winter when the longer coat is much lighter. *Source:* courtesy of Dyan Westvang.

thorough understanding of the genetic control of horse color can be helpful in such instances because an observer is then at least able to understand how the color arose, and this can help answer the question of color terminology if only through the back door. The overall process of correctly identifying horse color begins with the descriptive stage. Descriptive knowledge helps in the understanding and appreciation of the biologic (genetic) control of color. Finally, knowledge of genetic control circles back to enhance the appreciation of descriptive categories and subtleties.

Horse color is generally believed by most breeders to be only a cosmetic detail. But many horse owners, and indeed entire horse-using cultures, have long attributed specific characteristics to horses of specific colors. Such beliefs are generally dismissed as fiction, but a kernel of truth may well lie behind a few of them. A small collection of European research has verified that horses of specific colors do tend to react somewhat predictably in certain situations and that these reactions vary from color to color. A general trend seems to be that horses of darker colors are livelier than lighter ones, but the breeds in which this was determined were not noted, so the range of colors is likewise uncertain. While any behavioral connections to horse color are far from proven, these tidbits of folklore will be noted as the different colors are discussed. Even though relationships of behavior and color are speculative, they are an interesting part of the art of horse breeding and horse keeping. The interaction of color and behavior is treated briefly in Hemmer's (1990) *Domestication: The Decline of Environmental Appreciation*, listed in the Bibliography.

1.2 Basic Principles of Genetics, Genomics, and Molecular Biology

A thorough understanding of the genetic control of horse color takes study and hard work. The reward is an enhanced appreciation of horses and the biology behind their beauty, as well as more accurate descriptions of horses for registration purposes. An understanding of genetics is almost essential for breeders interested in producing specific colors of foals. An overview of some principles of genetics as well as some definitions of genetic terms is a good starting point for this endeavor.

A horse's color results from the interaction of several generally independent processes (or factors). About 14 of these have been documented to date, and account for nearly all variations in horse color. Most of these processes are known to involve a specific gene. Others have only been characterized visually, such as **mushroom**, but have not yet been characterized genetically. Interaction of so many factors unfortunately means that the control of color is inherently complicated; it cannot be made completely simple. The interaction of all of the factors results in the many different shades and types of horse colors.

For most colors, each specific combination of interacting components results in a unique color. A few colors are exceptions to this rule. The genetic basis of the colors, therefore, neatly explains most of the colors by accounting for the complex interactions that cause them. These interactions can be understood best if the basic factors are taken one at a time. In this way the number of complex interactions can be broken down into fewer key components, and the colors can be understood sequentially by adding the effects of each of the processes in turn. The genetic basis for most of the 14 processes controlling color has been documented by a variety of scientific studies, resulting in a fairly complete understanding of them. However, the theoretical basis for some of the processes is currently only an educated guess.

Genes are responsible for all biochemical processes that occur in living organisms. They are defined as units of heredity and they carry the information from one generation to the next by storing the information inside the nucleus of each cell. Genes code for the necessary information to make life processes happen, as explained in more detail later. Genes themselves are composed of DNA. DNA is made up of building blocks that are linked together to form a long chain. These building blocks are known as nucleotides, and they come in four types. For simplicity these are referred to by the abbreviations A (for adenine), T (for thymine), C (for cytosine), and G (for guanine). Each gene has a unique sequence of these four building blocks of DNA, allowing it to store the information to perform a specific function in the cell. This information typically codes for proteins. The stored information in the form of DNA (nucleic acid) can be translated into a new language for proteins (made of amino acids rather than nucleic acids). Proteins perform most of the functions in the cell, so the DNA code directly affects cell function.

It is the precise order and combination of the DNA nucleotide building blocks that makes one gene different from the next. Any given gene can have slight modifications in the order of the building blocks. These alternative forms of genes are known as alleles. Alleles differ from one another in many possible ways. The alleles are referred to as polymorphisms, meaning "many forms." A simple example is the situation in which one nucleotide differs between two alleles of a given gene. For example the short sequence of

nucleotides TAGACAT could be changed to TACACAT (genes have nucleotide sequences of hundreds or more than thousands of nucleotides, so this short sequence is not an entire gene). In this example "C" replaces the third "G." This difference is called a single nucleotide polymorphism, or SNP for short. This variation can be translated differently from DNA to protein, much the way "TAG A CAT" means something different than "TAC A CAT." Modifications in the order of the nucleotide building blocks change the function of the gene because the resulting protein has a different sequence of amino acids, all caused by changes in the original genetic code. When the genetic code is altered in this fashion the result is called a "missense mutation." This is a change in the DNA sequence that results in substituting one amino acid for another in the final protein product.

Other changes include insertions of extra nucleotides into the sequence, which shifts the way the code is read and therefore the final product. Another type of change occurs when portions of the code are deleted. The missing segments likewise change the final message, and therefore the final protein that is made from the message.

Most genes in organisms that have cells with a nucleus (including horses) are organized by alternating different sorts of DNA sequences. These are known as "exons" and "introns." The sequences that actually code for the protein product are the exons; as a result, the changes in those regions can change the final protein product, and therefore also change its biologic function. The exons are interrupted by the introns, which are sequences that do not code for the final protein. While the introns do not code for the protein product, many of the introns do have important regulatory functions that instruct the cell as to how to put the exons together. The consequence of this is that mutations in the intron sequences can still have consequences for the final action of the gene. Two examples, discussed in more detail later, are *brindle 1* (Chapter 3) and *sabino 1* (Chapter 7).

In horses, as in all mammals, genes occur on chromosomes. A simple way to imagine this is that chromosomes are long stretches of DNA. The DNA is folded up with proteins known as histones to fit inside the nucleus of the cell. Horses have two copies of every chromosome for a total of 32 pairs (Figure 1.7). Each pair of chromosomes contains specific genes.

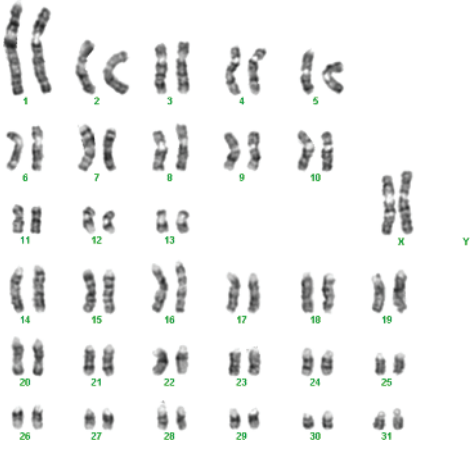

Figure 1.7 The microscopic chromosomes in a cell can be photographed, cut apart, and then arranged in their pairs. The resulting array is called a karyotype. Horses have 32 pairs of chromosomes. *Source:* courtesy of Dr. Teri Lear.

The complete set of an organism's genes is referred to as its genome. The genome represents the DNA from one set of the chromosomes. The first horse genome (from a **grey** Thoroughbred mare) was sequenced in 2009. An entire horse genome contains approximately 2.7 billion DNA nucleotide building blocks.

The most modern techniques reveal that chromosome organization is very complicated, but they can still be loosely thought of as strings of genes. The genes are something like beads of information, and thinking of them this way facilitates understanding the basics of their function as genetic units passing from generation to generation.

Chromosomes occur in pairs; an individual gets one of the pair from the sire and the other from the dam. When an individual reproduces, it contributes a random half of its chromosomes (one of each pair) to its offspring. The other half of the resulting offspring's genetic makeup comes from the mate. This halving of genetic material from parents and pairing in offspring is the mechanism by which the genetic code works its way through generations and populations, and is also the source of variation from parents to offspring as the combinations are reshuffled in each generation. This separation and rejoining is the basis of heredity and explains how two offspring from the same two parents can look different—they each received a different "half" of the genetic material from the parents.

It is essential to understand the concept of halving and subsequent rejoining into pairs at each generational step in order to appreciate the impact of genetics on horse color (Figure 1.8). The recombination of pairs at each generation is the basis of how genetics operates. The components of the pairs are constrained by the specific variants in the parents, and only those parental variants will be available for contribution to the offspring. Recombining of genes from generation to generation is frequently called "segregation" because the genes can appear in different combinations at each generational step.

Each gene occupies a specific site on a specific chromosome. This site is called a locus (plural, loci), and frequently genes are described by their locus names. Locus simply means an address for a gene; it is a specific physical place that a gene occupies. For example, the gene responsible for appaloosa patterns is located at a specific locus on horse chromosome 1 (Figure 1.9).

Both members of a pair of chromosomes in a horse have identical loci, all lined up in the same sequence. The only exceptions are a few, generally pathologic, situations. Usually, the term locus, even though singular, applies to a specific site on both chromosomes of an individual animal. For example, the *Agouti* locus in a horse implies the specific genetic site on two chromosomes (one from the sire, and one from the dam), each coding for the same piece of genetic information. Specifically, the *Agouti* locus is found at a specific site on horse chromosome 22. It is found at that same site on the chromosome 22 inherited from the sire as well as the chromosome 22 inherited from the dam, and indeed is at that same location in all horses (Figure 1.10).

When a gene occurs in more than one form (due to a change in the order of the nucleotides), the different forms are called alleles. For example, the *Agouti* locus has two well-characterized alleles (Figure 1.11). One allele, A^a, is missing 11 base pairs (building blocks) when compared with the other allele, A^A. The alleles of a gene all occur at the same locus, although each chromosome can only have one allele. The result is that each horse has, at most, a total of two different alleles at any locus because it has only two of each chromosome. The wide variety of horse colors results from individual horses

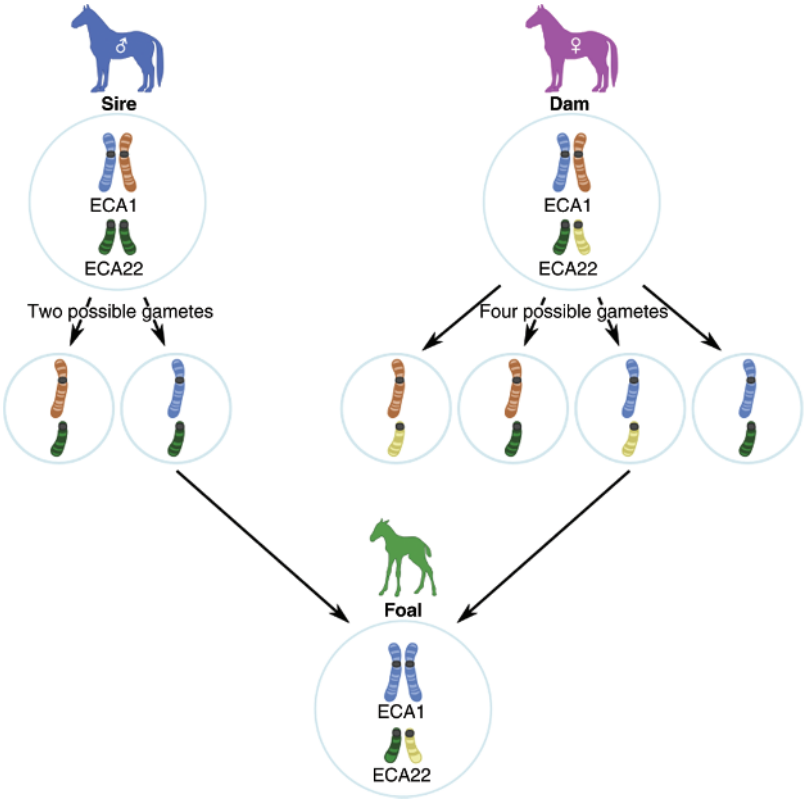

Figure 1.8 Pairs of chromosomes separate at random during the formation of egg and sperm, and these pairs then rejoin during fertilization. This separation and rejoining is the basis of heredity. This figure shows only one possible result, and illustrates how new combinations can be formed as a consequence of the process of reshuffling the chromosomes. *Source:* courtesy of Francesca Gianino.

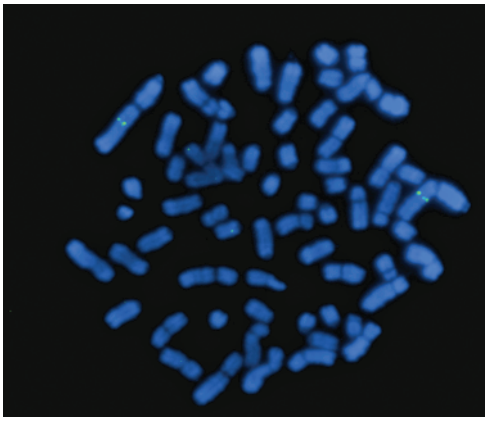

Figure 1.9 This is a horse karyotype as it appears before cutting the image and pairing up the chromosomes. The green dye indicates the locus for appaloosa pattern (*LP*), and lights up the same specific site on both copies of chromosome 1. *Source:* courtesy of Dr. Teri Lear.

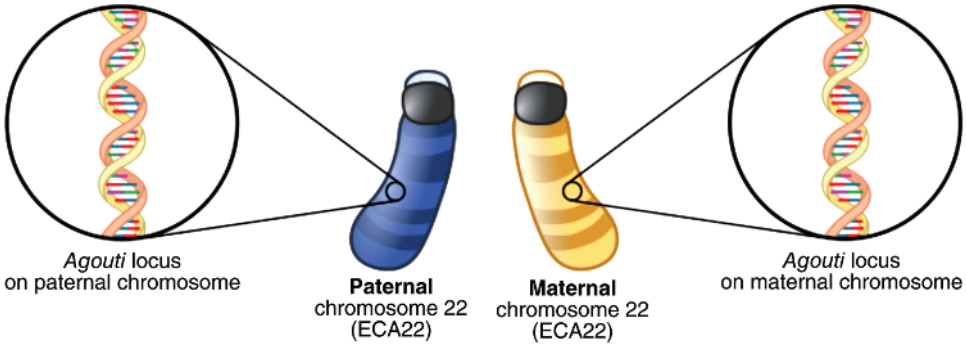

Figure 1.10 This is a schematic of chromosome 22, highlighting the site of the *Agouti* locus on both the maternal and paternal chromosomes. The *Agouti* gene is found at this location in every horse. *Source:* courtesy of Francesca Gianino.

differing from one another in the specific allelic combinations at the many loci controlling the different components of color. These numerous allelic combinations cause the variations that humans perceive as the range of colors in horses.

The specific combination of alleles, or genetic makeup of a horse, is called its genotype. The external appearance or physical makeup is called the phenotype. The phenotype may or may not completely reveal the underlying genotype. "Genotype" is used for the specific genetic variants that are present in an animal, whether at one locus or many. This is subtly different than "genome," which is the entire DNA within an animal.

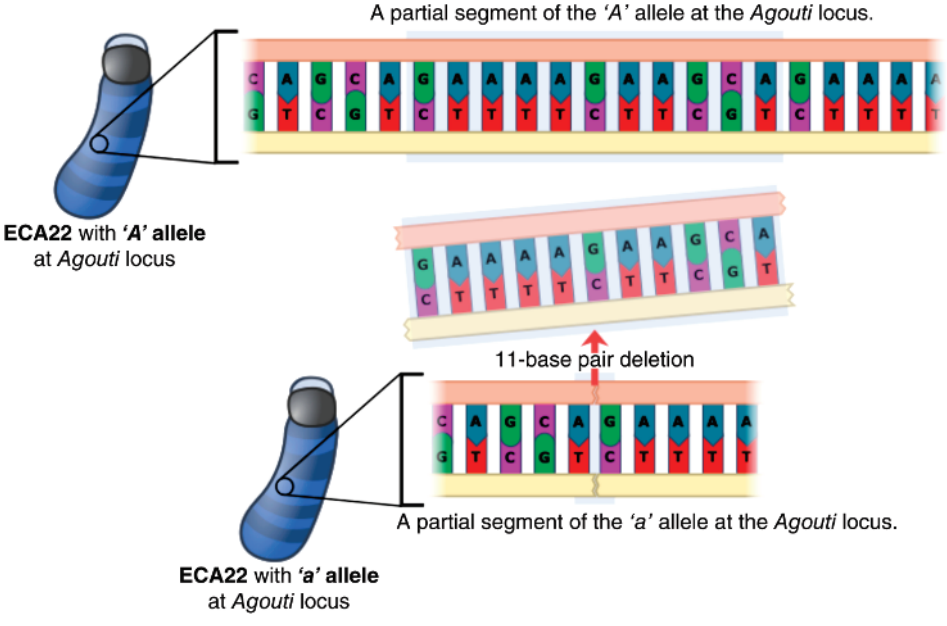

Figure 1.11 The two known alleles at the *Agouti* locus. *Source:* courtesy of Francesca Gianino.

The condition of having two identical alleles at a locus (one on each chromosome) is called "homozygous." When the alleles are different, the situation is called "heterozygous." This terminology (homozygous/heterozygous at a locus) reflects back to the concept of "locus" as encompassing a specific site on both of two chromosomes that codes for the same gene and the fact that an individual horse only has two of each chromosome. As a result, the only options at a locus are to have two identical copies of the gene (homozygous) or one copy of each of two different alleles of the gene (heterozygous).

Alleles at a genetic locus interact in a variety of ways. Some alleles are not expressed phenotypically unless both copies of the gene in an individual are the same (homozygous). These are recessive alleles (or genes, the two terms are sometimes used interchangeably). Dominant alleles, in contrast, are expressed identically whether in one copy (heterozygous) or two copies (homozygous). The main concept is that a dominant allele masks the phenotypic expression of a recessive allele when the two are paired together in the heterozygous condition. As a result, colors associated with recessive alleles can appear to pop up out of nowhere because they can be carried along, unexpressed phenotypically, as long as they are paired with a dominant allele. They can be carried along this way for several generations.

Recessive alleles are perceived as surprises when they become expressed by virtue of being paired in the homozygous offspring of two individuals that carry them. In this case, the carrier parents do not show the effect of these alleles because they are being masked by a dominant allele. Dominant alleles cannot be carried along in a hidden state in this manner. If a dominant allele is present, it is expressed phenotypically. As a result, the presence of dominant alleles shows up in each generation and they are therefore are not the source of surprises except very rarely, such as in the case of a spontaneous mutation. Mating of recessive to recessive can yield no surprises because nothing is hidden, and the phenotype therefore betrays the genotype completely, or nearly so.

It is important to understand that the character of an allele as dominant or recessive is inherent in the functional capability of the allele itself. This does not change over time, nor does it change in various situations. Many people are under the mistaken impression that dominant alleles are necessarily common and that recessive alleles are necessarily rare. The issue of allele frequency is totally separate from the issue of dominance and recessiveness. Some recessive alleles, such as *chestnut*, are indeed common to the point of being uniform throughout some breeds of horses, such as the Suffolk or the Haflinger. The uniformity of the *chestnut* allele in the Suffolk in no way changes its character as a recessive allele, so that crosses of Suffolks to black Percherons result in few if any chestnut foals, because the black points of the Percheron will dominate the *chestnut* allele of the Suffolks. Likewise, some dominant alleles, such as *white*, are incredibly rare or nonexistent in most breeds and yet are routinely passed as dominant alleles in those few families in which they occur.

Some alleles are described as incompletely dominant, which means that two, one, or no copies each results in a separate appearance. Each situation can be detected phenotypically by examining external appearance. Incompletely dominant systems are the easiest to understand, because no surprises can result, such as occur with hidden recessive alleles. With incompletely dominant alleles the two different homozygous genotypes, as well as the heterozygous genotype, each have a distinct phenotypic appearance.

Another interaction of genes is called epistasis. This refers to the ability of specific allelic combinations at one locus to mask the expression of alleles at another locus. Epistasis is another example of the complexity of genetic interactions. It is similar to the relationships of dominant and recessive alleles, but concerns two or more loci instead of only one. The locus that is masked by an epistatic gene (or allelic combination) is referred to as being hypostatic, while the locus or gene combination causing the masking is called epistatic. Hypostatic genes can pop up as surprises, much as do recessive genes when masked by dominant ones. Specific examples in the discussion of the colors will illustrate this phenomenon and will help in the understanding of what can be a subtle and confusing concept.

Genetic loci can be considered as separate little biochemical factories. Each locus controls some unique aspect of the color that is finally produced and seen. It is convenient to consider the control exercised by each locus as a decision mechanism. At most loci the choice is either "situation A" or "situation B." Thinking about the loci as responsible for separate decisions helps in understanding that each locus presents a choice, and the sum of all the choices will affect the resulting color. If each locus is considered to make a separate component of the final appearance, it is easier to understand how color arises as the final result of specific combinations of choices.

As an illustration, consider horse color to be like a cup of coffee. The first choice is to have dark roast or light roast, which determines the base color. The next choice is to add cream, or not, which will either dilute the base color or leave it dark. Finally, any sweeteners can be added, just like adding white patterns to a horse. Much in the same way that the final cup of coffee is dependent on which choices were made at each of the decision steps, a horse's final appearance is the result of the combination of all the options encoded in the alleles of the genes. By understanding the components it is possible to appreciate the interactions that led to the color that is present on the horse. Because each locus has only two (or a few) choices, the components are fairly easy to understand once they are identified and appreciated as basic components of a final combination. The concept of each locus doing a separate job is the key to understanding the genetic basis of horse color. By viewing genetics in this way it is possible to appreciate how the various colors can be built in successive steps from the various component parts.

Genetic nomenclature is subject to a number of conventions, and these are variable in much of the literature. A standard format is used throughout this guide, and follows the guidelines of the Committee on Genetic Nomenclature of Sheep and Goats, which has been expanded to now include other domesticated animals. Names of loci have an initial capital letter to distinguish them from names of alleles, which are all in lowercase letters. Both loci and alleles are in italic type. *Dun* is a locus, *black* is an allele. The symbols for loci are abbreviations of the names and are in italics, with an initial capital letter. Allele symbols are added to locus symbols as superscripts to separate them from the locus symbol. The symbol "+" is used to denote the probable wild-type allele at a locus.

For example, Dn^{+} is the *wild-type* allele at the *Dun* locus. The wild type is inferred from the probable original color of horses before their domestication. Inferring the wild-type color is important for genetic nomenclature, but has become more difficult because the sequencing of DNA from fossil remains has indicated multiple color variants were present in horses before the time of domestication. While the list is incomplete,

Figure 1.12 This light **zebra dun** with the **mealy** overlay is likely the original color of horses. *Source:* courtesy of Jeannette Beranger.

Siberian Pleistocene animals included **bay**-based horses and a few others. Some evidence points to the *leopard* allele being present before domestication. Similarly, the *black* allele and the *non-dun* allele have also been found in at least some pre-domesticated horses. While assigning a single wild-type color is increasingly difficult, it is most likely that a light yellow, shaded version of **dun** was the most common wild-type color, similar to the color of today's Przewalski's horse and present only rarely in most horse breeds (Figure 1.12). This color is taken as the single wild-type color in this work, despite its limitations.

Symbols for alleles other than *wild type* are standardized so that dominant alleles have an initial capital letter, while those for recessive alleles have a lowercase letter. This convention is used even though names of both dominant and recessive alleles, when spelled out and not referred to by abbreviated symbols, have an initial lowercase letter. As a result, *dominant black* is an allele proposed at the *Extension* locus, and has the symbol E^D.

Another convention is the abbreviation of genotypes by using dashes to fill in behind dominant alleles when the second of the pair is unknown or unimportant for determining phenotype. For example, A^A— (*bay* at the *Agouti* locus) can be used for A^AA^A and A^AA^a, which both appear the same phenotypically. A^aA^a (a homozygous genotype with two *black* alleles at the *Agouti* locus) has no abbreviation, because a recessive genotype masks nothing, although sometimes a single A^a is used as a phenotypic abbreviation for A^aA^a.

A further problem in nomenclature is becoming increasingly frequent as the biochemistry and molecular mechanisms of the gene products become established. It is now

common for the older symbols to not match the loci documented by results of molecular investigations. Most of the newer locus names and symbols are based on mouse nomenclature, because the molecular genetics of mice has been so much more extensively characterized than has that of any other species. For example, the *Agouti* locus encodes for the protein called the agouti signaling protein (ASIP). The result is that this locus, traditionally represented as the *A* locus, is also represented as the *ASIP* locus in modern studies. In this text the molecular loci are noted, while the older symbols are still retained because long use has made these widely accepted and understood. To change them in order to match the molecular aspects would increase confusion, rather than reduce it, because so many of those interested in horse color first encountered the earlier abbreviations and are now familiar with them.

Another convention used throughout this guide is the printing of names of horse colors in bold type. This reduces confusion between a discussion of a color in general and a specific name for a horse color. Black, for example, is a general name for a color in nature, while **black** is specifically the color of a **black** horse. Other peculiarities of eye color and hair are also printed in bold type because these also have specific connotations when their use concerns horses.

1.3 Pigment Cell Function and Genetic Control

A fundamental principle to understand about horse color is that color in horses is due to the presence of pigments in the hair. Two major pigments account for all colors of mammals (including horses). One of these is eumelanin (YOO mel a nin), which is responsible for black or slate blue. In a very small number of horses eumelanin is brown (flat, chocolate brown) rather than black. This chocolate brown type is similar to the color that is common in retriever or spaniel dogs. The switch between black and brown eumelanin is an "either/or" phenomenon for the whole horse: a horse can form either black or brown, but not both. This is an important detail, even though the colors based on brown are so rare.

The other pigment is pheomelanin (FEE oh mel a nin), which produces colors ranging from reddish brown or tan to yellow. Pheomelanin can vary in shade on a single horse, and on many individual horses pheomelanic areas do indeed vary from dark to light. Most horses have both pheomelanic and eumelanic areas on them and are therefore combinations of black along with red- or yellow-pigmented areas. Dark pheomelanin sometimes can resemble brown eumelanin and is much more common than eumelanin as a source of any brown areas on horses. Pheomelanin usually retains at least some of its reddish color, even when very dark, and this reddish tinge helps to distinguish pheomelanic areas from eumelanic areas.

White hairs result from a lack of pigment granules, and these are essentially hairs without color. Skin that lacks pigment granules is characteristically pink, and it gets its pink tone from the presence of blood in small, superficial blood vessels.

Color in horses is possible because pigment cells, called melanocytes, act to put pigment granules into cells that become hair and skin. The presence and function of these cells determines the amount, type, and character of pigmentation. Melanocytes migrate to the skin in embryonic life. They originate along the neural crest, a specific group of embryonic cells that also give rise to the spinal cord and brain. The importance

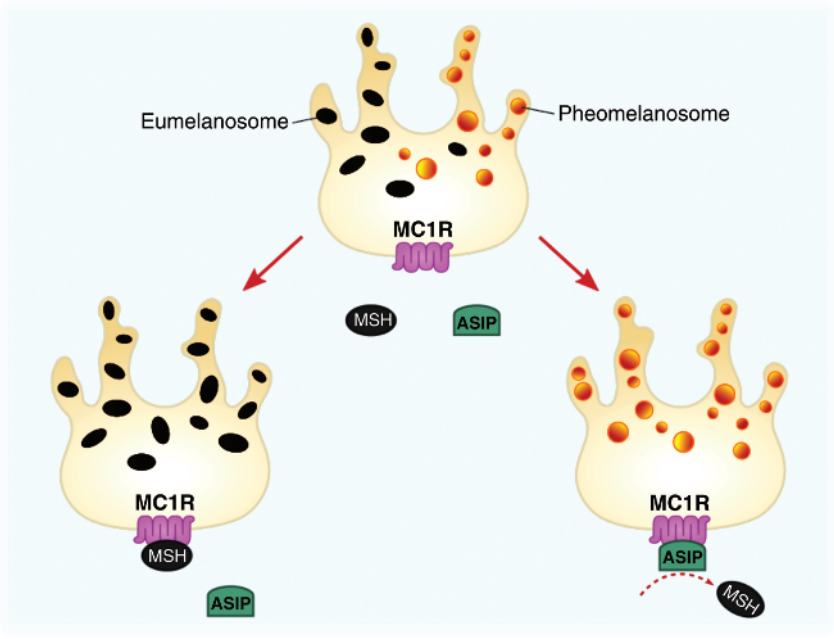

Figure 1.13 Pigment cells are capable of producing two types of pigment, which is regulated by the activation or inactivation of a cell surface receptor (MC1R). When MSH is bound to the receptor, it activates the receptor and the cell then produces eumelanin (black or brown pigment). When ASIP is bound to the receptor, then the receptor is blocked, cannot be activated, and the cell produces only pheomelanin (red or yellow pigment). *Source:* courtesy of Francesca Gianino.

of this detail is that the pigmentary system and the nervous system are closely allied in embryologic life, and some specific genes affect both of these systems instead of only one or the other.

Melanocytes can produce either pheomelanin or eumelanin (Figure 1.13). The determination of which pigment is formed is accomplished by a receptor that is on the surface of melanocytes. This receptor is the melanocortin 1 receptor, or MC1R for short, and is activated by a protein known as melanocyte-stimulating hormone (MSH). This hormone is produced by the pituitary gland, located at the base of the brain. In the absence of activation by MSH, cells form pheomelanin. When the receptor is activated, the result is eumelanin formation.

The switch between eumelanin and pheomelanin production can be influenced at several different steps. One switch is the presence or absence of MSH (the protein hormone that activates the receptor just described). This is a fairly rare switching mechanism in animals, and is unimportant in horse color because MSH is thought to be constantly available to all cells of horses.

A second switch is at the level of surface receptors. These are produced by genetic codes at the *Extension* locus as well as a few other loci. In other words, the *Extension* locus codes for MC1R. Some mutations, such as *dominant black*, produce a receptor stuck in the "on" position in many species. This is called a gain-of-function mutation. The result is an entirely eumelanic, usually black, phenotype. In this situation the

receptor is always activated, even in the absence of MSH. Other mutations, such as *chestnut*, result in a totally inactive receptor (this is called a loss-of-function mutation) that is incapable of responding to MSH and therefore results in a completely pheomelanic (red/yellow) phenotype in most cases.

A third way to affect the switch is to block the surface receptor externally. This leads to an inability of the receptor to be activated even though the receptor is normal and MSH is present. This mechanism is typical of the *Agouti* locus and results in pheomelanin in those areas of the body that express the agouti protein (causing a blocked receptor) and eumelanin in others where this protein is lacking (unblocked receptor). The regional distribution of the agouti protein is under genetic control.

The internal workings of the melanocyte are as important as the surface receptor in determining the final production of pigment. The formation of eumelanin and pheomelanin involves several steps, some of which are common to both pigments and some of which are unique to only one pigment or the other. This is important because some mutations affect the production of only one of the two pigments, while others affect the production and character of both. The loci controlling the internal packaging and production of pigment in horses are increasingly being characterized at the molecular level. As a result, past uncertainty concerning homologous loci (those with identical function) in other species is slowly being resolved. This contributes to deeper understanding of the specific mechanisms leading to the various colors of horses.

The pigments are formed in small packages called melanosomes. These can be moved from the melanocyte to surrounding hair and skin cells. The packaging and distribution of melanosomes is another potential site of variation, which can result in different visual appearances of colors that are caused by identical pigments. These changes are under genetic control at yet other loci distinct from those already mentioned.

Mutations can occur in the surface receptors themselves (*Extension* locus), or the proteins that bind directly to them (*Agouti* locus). The consequence of these surface changes is stimulation or blockage of these receptors, which results in changes in the type of pigment produced (eumelanin versus pheomelanin). In contrast to these surface changes, changes in the internal machinery of the melanocyte usually result in changes in the amount of pigment that is produced or how it is packaged. The consequences of internal changes usually then determine the degree of dilution of pigment rather than changes in the specific pigment type. These include loci controlling *Dun*, *MATP* (*Cream and Pearl*), *Champagne*, *PMEL17* (*Silver Dapple*), and *Mushroom*. The difference between the two sites ("surface" versus "internal to the cell") is basically one of "which" (surface) versus "how much" (internal).

White patterns are controlled at yet other loci that affect the migration, survival, or function of melanocytes. A handful of loci consistently cause white patterns in several species. One of these is the receptor for endothelin B, which is affected by the *frame* mutation. Other good candidates for white-pattern mutations include the mast cell growth factor locus (the *KIT* locus). This locus appears to be affected by many mutations and is thought to be the locus responsible for several different white patterns such as *roan*, various *sabino* mutations, and *dominant white*, among others. This locus is also very close to the allele for *tobiano*. The homologies of white-pattern alleles with those of other species are still not fully documented in horses.

By mid-2016, 40 polymorphisms in at least 14 genes had been identified that contribute to or are associated with horse color variation. While the resulting

biochemical mechanisms have been extensively studied for some of these, the functions of others are only speculative. Some of these discoveries were aided by the sequencing of the reference horse genome that was published in 2009. Additionally, sequencing the genomes of horses from a variety of different breeds has allowed for the development of new tools that help to identify chromosomal regions and mutations associated with different phenotypes. These include coat color as well as disease and performance traits. These new tools, along with new sequencing techniques, will undoubtedly hasten the pace of characterization of the genes involved in different color phenotypes, and will also reveal novel genetic mechanisms involved in producing them.