

Anderson *et al*., 2020

**The Chicken *E* Locus Nomenclature**

Ronald Okimoto

11/14/2024

My first blog post is going to be about the figure on the blog’s home page. It is a cartoon depiction of the **phenotypes** associated with the chicken *E* **locus** (extension of black) that I put in a book chapter on feather color **genetics** (Andersson et al., 2020). These **alleles** were named before we knew what **genes** were. I will use the figure to compare the classic genetic nomenclature to the current gene designation nomenclature that started to be used in the 1990’s. What I hope to accomplish is to provide a direct and simplified description of the past and current nomenclature, and while doing so introduce the readers to the genetic vocabulary that they will need to understand future explorations into genetics.

The classic nomenclature had allele and locus designations. The *E* locus had extended black, birchen black, dominant wheaten, wild-type, brown, and recessive wheaten (*E*, *ER*, *eWh*, *e+*, *eb*, and *ey*, respectively). There were other *E* locus alleles that had claimed to have been identified, but they were not very well characterized. I recommend Bob Smyth’s chapter in Poultry Breeding and Genetics, 1990. Edited by Crawford for a good review of what was and is known about feather color genetics because we really haven’t invested very much research hours into the subject except to try and identify the genes responsible for the feather color phenotypes, and the phenotyping correlating feather color with the *E* locus for the last couple of decades hasn’t been well executed in terms of trying to figure out what alleles were segregating.

You can have **dominant** alleles whose phenotypes are expressed over the phenotype of the allele that is **recessive**. The phenotype of an individual with an *E/E* **homozygous** **genotype** is fully black feathered, the *eb/eb* individual has the wild-type black breasted red phenotype in the male and brown in the body of the female, and the *E/eb* **heterozygous** **hybrid** is nearly fully black for males and females with that genotype. So extended black (*E*) is dominant over the brown (*eb*) allele. Alleles with dominance were usually designated with capital letters while recessive alleles were designated with lower case letters. I am likely a member of the last generation of geneticists that had to suffer through typing these designations on a typewriter. You had to change the ball on a Selectric or underline the letters that should be italicized, and roll down to type the superscript. Word processors came in when I was a graduate student, but early programs like Word Perfect were not much better with their embedded print commands.

The current gene nomenclature has the gene name and then the allele designation. Many of the initial new allele designations using the current nomenclature will likely have to be changed in future literature. **DNA** sequence variants are often given different allele designations as numbers because it is unknown what classic allele the sequence variant may represent. This was done for many of the initial *E* locus *MC1R* gene sequences. It is problematic to relate the classic allele designations to variant numbers. For the figure I and my coauthors tried to use allele designations that could be related to the classic designations. You have the gene name an \* and then the allele designation. I did not put dominant wheaten in the figure (it has the same adult phenotype as recessive wheaten) because we haven’t figured out which wheaten allele has been identified in our sequence analyses. So in their order in the figure the current allele designations are *MC1R\*E*, *MC1R\*R*, *MC1R\*N*, *MC1R\*B*, and *MC1R\*Y*. No lower case is used to identify the more recessive alleles and *N* (for normal) replaces the + (**wild-type**) designation.

The current nomenclature makes It difficult to relate to the older literature that first determined the genetics of the phenotypes. For this reason, when the older literature is needed to clarify current research, I will likely be using the classic nomenclature used in the older publications.

**The vocabulary of Genetics**

Ronald Okimoto

11/14/2024

**Jargon**: special words or expressions that are used by a particular profession or group and are difficult for others to understand.

Definitions from [Oxford Languages](https://languages.oup.com/google-dictionary-en)

The following section can be referred to if you do not understand genetic jargon. It may also be a way to take a short course in basic genetics to review your existing knowledge and possibly understand what you may have been missing. I do not go deeply into theory or the math, I just explain in general terms what is meant by the jargon term under discussion.

When I taught Genetics at the University of Arkansas I would tell the students in the first lecture that understanding a scientific subject like Genetics required learning a new language. Scientists invent a vocabulary that you need to deal with in order to understand what they are talking about. We do not do this to make Science some sort of exclusive club, but it is done because we are lazy communicators. Writing and speaking on a subject is a major part of the scientific endeavor, so each field develops a shorthand type of vocabulary that allows us to put a paragraph of explanations into one sentence. If I had to explain what a “gene” or an “allele” was every time those terms came up in a conversation or in a publication I would go nuts. As a graduate student my communication skills were so bad that my committee members made me take a technical writing course and I took more than twice as many seminar classes as required. English is not my second language. My father was born in California, and my mother was born in British Columbia, and I was born and raised in the USA. I just had terrible language skills.

In order to help students that might be as language impaired as I am, I would give a list of terms with the lecture outline, and would give a weekly quiz on them so that the students could earn some easy points by defining the vocabulary list. I warned the students that if they fell behind in understanding the scientific jargon of genetics that they would encounter increasing difficulty in understanding each lecture and the reading going deeper into the textbook. In the nomenclature post I have indicated some terms in **bold** that the reader would likely need to understand, and defining that jargon has issues with including terms that also have to be defined. Below I have tried to define the terminology that is needed to understand how the simplistic initial post relates to the study of genetics. What occurs is that jargon terms appear in the context of other jargon descriptions, so the reader gets to see the term in context and can scroll around to get a better description of what the term means. It may be a good way to introduce the jargon into your own vocabulary. It is a short course in genetics.

**Additive and non-additive genetics**: Since this is the first alphabetical entry I am going to take the opportunity to relate basic biology to genetics. In the beginning we did not know the biological basis for inheritance. Mendel knew that he had to take pollen from one plant and dust it onto the stamen of another plant to make his crosses, but he did not know what was happening in terms of what the genetic material was, nor how the genetic material was transferred and formed the **diploid zygote** by the fusion of pollen and egg nuclei. Geneticists had to rely on **phenotype** (the physical attributes of the lifeform). Mendel and the geneticists that followed him estimated the genetic component by the variance of the phenotype. Essentially the phenotype was used to track the inheritance of the genetic factors. Even today when we have **genome** sequences we still have to rely on the phenotype because we do not fully understand the genome, do not know what all the genes do, and what is responsible for most of the regulation of those genes. So we continue to rely on the phenotype to estimate the genetic components. When I am discussing the genetic component I am usually referring to the effect on phenotype and how the genetic variance of a trait relates to the portion of the phenotypic variance explained by that genetic component.

<https://en.wikipedia.org/wiki/Genetic_variance>

The phenotype of the organism is not totally due to the genetic material that it inherited from it’s parents. Life comes from life, so what you are is dependent on what came before. The sperm may be basically a DNA delivery system, but the egg cell preserves the biology necessary for the production of new life passed down through the generations. The egg cell contains everything needed to make new cells. You can’t just put sperm and egg nuclei into a lipid bilayer bag of water and produce a new human. A lot of what you are is due to the long lineage of cells going back to whatever the first lifeforms were, probably, before the **genetic code** evolved.

There is a paternal and maternal component to phenotype and some of this maternal and paternal component has a genetic basis. A maternal component may be where the placenta attaches in the uterus. Twins may have their own placentas and one twin may get short changed in position of the placenta and it can have a negative effect on the affected twin, and the twins can have very different birth weights. The womb environment may influence the individuals growth and development after birth. Placental placement is likely due to chance and not some genetics that the mother has, so it is an environmental influence. After birth the mothers may have different milk supply characteristics. The milk can vary in protein, fat, and sugar content as well as in quantity. This maternal effect does have a genetic component, and the mother’s progeny are expected to not only be affected by the quality and quantity of the milk, but they could have also inherited some of the genetic component for this aspect of their environment. The environment has a large effect on phenotype. We use the equation P = G + E (Phenotype is genetics + environment). Some of the paternal and maternal influence can go into genetics and some is dealt with as environmental.

**Additive and non-additive genetics** are **quantitative genetic** terms that relate to how multiple **alleles** at multiple **loci** each contribute a bit to a **quantitative** **trait**. Combined the additive and non-additive genetics are given a numerical value as **heritability in the broad sense** (**H2**). The additive component of a quantitative trait such as body weight is given a numerical value as **heritability** **in the narrow sense** (**h2**). The higher the h2 value for a trait the larger is the additive genetic contribution for that trait. The h2 estimate of an individual reflects the parts of that individuals genome that contribute plus or minus to a quantitative trait. Each variant **locus** affecting the trait adds or subtracts a bit from the trait if it is inherited by the progeny of that individual. Breeders like to see high h2 values for the traits that they are breeding for because the additive component is the most reliable genetic component for estimating the value of individuals as breeders. The higher the h2 the more like the **parent** the progeny are expected to be. So when we “breed the best to the best and hope for the best” breeders want the h2 estimate for the trait to be as high as possible. The percent breast meat on the chicken that you eat has an h2 of around 0.4, body weight has a h2 of around 0.2, and a trait like mortality during lay may have an h2 less than 0.1.

One reason that traits are not fully heritable is because the environment has a lot to do with expressing the full genetic potential. The environment can have very significant effects on a trait. A disease may infect a flock and many birds may die irrespective of what livability genetics they may have. Individuals may suffer famine, and never grow to their genetic potential. There are also non-additive genetics affecting most quantitative traits that reduce the proportion of the additive genetic effects. Non-additive genetic effects are due to things like **gene interactions** that produce a phenotype or **dominance** observed among the **alleles** of a gene **segregating** in the population.

Dominance is likely the easiest to understand as to how it affects h2 estimates. For breeding it is important to understand that the allele frequency affects the accuracy of selection as well as the ability to estimate heritability. Recessive white (*c*) is an allele of a single locus (*C* locus) with a large **epistatic** effect on feather color. If you have a flock of black feathered chickens and introduce one bird with recessive white. White feathers has a heritability of zero in that first cross. None of the progeny are white feathered. If you did not keep a **pedigree** you would not know which birds must have inherited the recessive white allele. White feathers will not reappear in the black feathered flock until you mate birds together that share the white feathered bird as an ancestor. Unless you **inbreed** (mate sibs in that first **hybrid** generation) you do not produce white feathered birds and h2 for white feathers remains zero until related individuals begin to be mated together. Serious breeders keep a pedigree and they would know that all progeny of the hybrid cross are **heterozygous** (*C+*/*c*), but the heritability of the first **backcross** to the black feathered line is still zero if the breeder does not want to breed relatives. None of the progeny will be white feathered, and only half of the backcross progeny will have inherited the *c* allele. A breeder can resort to inbreeding to reveal **carriers** of the **recessive** *c* allele, but inbreeding has negative consequences, so commercial breeders usually resort to **progeny testing**, and mate the possible carriers to a recessive white line to identify the carriers (heterozygotes will produce half white feathered progeny in the test mating). This slows down **selection** of the next generation of breeders, and now with the advent of molecular marker testing we can identify carriers using DNA testing. The point is in order to avoid inbreeding you have to suffer through several generations where heritability of white feathers is zero, and you have to track carriers by some other method than using the **phenotype**.

You can start producing birds with white feathers by mating cousins without suffering large amounts of inbreeding, but these white feathered birds will continue to have a low heritability for white feathers because the **allele frequency** of the *c* allele is low at this point of **introgressing** the *c* allele into the black feathered population, and it remains unlikely that a white feathered bird produced by cousin matings would mate with a heterozygous carrier and have a chance of producing white feathered progeny. It isn’t until the frequency of carriers becomes significant in the population that you start observing a measurable h2 for the white feather trait. As an example a white feathered bird’s progeny are less likely to be white feathered if the *c* allele has a frequency of 0.05 than if the frequency was 0.5. If the allele is at **Hardy-Weinberg equilibrium** for the *C* locus at an allele frequency of 0.05 very few of the birds will be white feathered under random mating (2.5 per 1,000, 0.25% of the population) and 9.5 out of 100 (9.5% of the population) would be heterozygous carriers), so few matings involving white feathered birds would be expected to produce white feathered progeny. Once the allele frequency reaches 0.5 you can have 25% of the flock with white feathers, and half (50%) of the flock will be carriers so matings with white feathered birds are much more likely to produce white feathered progeny and the heritability (additive component) for the trait can be estimated by phenotype of the parents and resulting progeny. So the additivity of an allele can be frequency dependent. For a recessive allele the additive genetic component in a population increases as the allele frequency increases in the population.

For a dominant trait like *C+C+* if you cross black feathered birds to a recessive white line, black feathers is highly additive in the first few generations. All the progeny in the first mating are black feathered. If you do not mate close relatives only half the progeny of the black feathered hybrids (*C+c*) will be black feathered when mated to *c*/*c* homozygotes, so the additive estimate goes down but remains relatively high, and stays at the lower level until you start mating black feathered birds together and ¾ of their progeny are black feathered instead of ½. It may seem odd, but for a fully dominant trait the heritability for that trait may never go to 1 in a population (all progeny may never be black feathered, but a rare white bird may appear every once in a while). Due to the vagaries of random chance and genetic drift once the allele frequency of the recessive allele drops to around 0.03 selection against the white feathered birds becomes inadequate to remove the allele from the population. You have to get lucky and not breed carriers in order to completely remove the allele from the population. This is why white feathered “sports” showed up in Barred Plymouth Rock lines and could be selected for to produce one of the original White Plymouth Rock lines. We have since found out that the White Plymouth Rock lines had the recessive white *c* allele of the Silkie breed and we know the allele must have come from the Asian breeds that went into creating the American dual purpose Standard breeds.

A side note may be of current interest. There was an embarrassing phase of **Eugenics** in American genetics whose greatest driving force was likely racism among the general population. Some well known early geneticists got involved with the Eugenics movement, but other geneticists essentially canceled the movement because most of the traits that the Eugenicists were willing to claim as being what they wanted to cleanse from the population were recessive deleterious traits. There is something called the **lethal load** or genetic load carried by individuals of a population. Humans have a lethal load of around 1.5 to 2. The numbers refer to lethal equivalents of recessive lethals in an individuals genome. Everybody has enough lethal alleles to kill themselves about twice over if the alleles were homozygous. This means that pretty much everyone is a target for genetic cleansing. It also could be shown that you couldn’t get rid of recessive deleterious alleles by just not allowing the affected to breed. You would have to resort to **family selection** and enforce the breeding limitations onto related individuals that had a good chance of not being **carriers** of the recessive deleterious alleles. Good sense eventually prevailed and the movement died. The concern today is that we have the technology to identify all of the carriers, and this would make the Eugenic goals obtainable. We only have our moral sense to stop a new Eugenics movement.

**Allele**: the original Mendelian definition of allele for a **diploid** organism defined allele as being one copy of a **gene** or **Mendelian factor**. A diploid organism like peas or humans have 2 alleles of the same gene or Mendelian factor. The alleles that an individual had could be the same or different. Mendel worked with only two alleles for each of his Mendelian factors such as yellow or green seeds, but genes can have multiple alleles like the *E* locus. Even though you can have more than two alleles **segregating** in a population any diploid individual can usually only have 2 alleles of each gene. At the beginning of genetics we did not know what a gene was, but we came to understand that the Mendelian factors existed on **chromosomes**, and that chromosome inheritance was consistent with Mendelian inheritance. The genes could be **linkage mapped** on a chromosome using **genetic recombination** analysis. Genes had to be specific bits of chromosomes. Variant alleles could be mapped to a specific **locus** on a chromosome, and it was this locus that contained the variant alleles. The literature is filled with mention of alleles of the *E* locus before we figure out that it was the melanocortin 1 receptor (*MC1R*) gene.

**Allele frequency**: if more than one allele exists for a **locus** in the population then each allele will account for a fraction of the total alleles found in the population. For two alleles **segregating** in a population you have the most common allele that is designated as the major allele. The less common allele is called the minor allele. The allele frequencies need to add up to 1, so for two alleles the major allele might have a frequency of 0.7 and the minor allele would have an allele frequency of 0.3 (0.7 + 0.3 = 1). Because we geneticists are usually most interested in the minor allele frequency it has the acronym designation of **MAF**, but we do not need a similar designation for the major allele frequency. If an locus has a MAF of less than 5% we do not use that locus for **genotyping** populations under **genomic selection** or for mapping **GWAS** (genome wide association studies) populations developed to genetically map specific **traits**. Not enough individuals of the population are segregating for the locus and using it doesn’t produce enough linkage information and is more likely to provide some spurious associations. Genetic **variants** with a MAF of less than 0.1% may not be counted in considering the **standing genetic variation** in a population. Basically, if you know the MAF you also know the major allele frequency, and we have more discussions about the minor allele frequency so it gets the “M” designation. These allele frequencies exist as genotypes in a population of **diploid** individuals, and the frequency of **genotypes** can be estimated from allele frequencies using the **Hardy-Weinberg equation**.

**Assortative matings**: is when there is a **selection** of breeders and there has been selection for specific genetics **segregating** in the population. No selection or random matings would be **non-assortative**.

**Backcross**: is a particular cross involving progeny and descendants of a cross between two **parental** lines. When an **F1** generation is produced by crossing two parental lines, if you take the F1 individuals and cross them back to either parental line the cross is called a backcross. For genetic evaluations the backcross is used to expose **recessive** **alleles** that came from the parental lines. The recessive **genotype** and **phenotype** segregates 1:1 in a backcross. Continued backcrossing is often used to **introgress** a particular **trait** into the parental line used as the recurrent parent. The F1 **hybrids** have one **genome** equivalent from each of the two parental lines. The F1 is usually claimed to have 0.5 genetics of one parental line and 0.5 genetics of the other parental line. The first backcross progeny retain, on average, ½ of the genome that came from the non-recurrent parental line, and the first backcross individuals are usually referred to as being 75% of the genetics of one line and 25% the other. This is true for individuals, but if you started with two highly inbred lines what you can determine is that each first backcross individual retains on average ½ of that inbred non-recurrent genome, so it is true that individuals are only 25% the genetics of one line, but each individual has inherited a different half of the genome from the non-recurrent parental line. As a population a lot of genetics is retained because each backcross individual has inherited a different half of the genome of the parental line not used in the backcross. You can recover over 95% of the genome of the non-recurrent parent among just 5 first backcross individuals. So if you use multiple individuals to create the first backcross and subsequent backcross you are not making the genetic progress that you might think that you are making. **Recombinant inbred mapping lines** depend on individuals retaining different portions of the genome not of the recurrent parent in the backcross.

**BLUP**: Is a statistical approach that is used as a genetic predictor for selection of breeders. <https://en.wikipedia.org/wiki/Best_linear_unbiased_prediction>

**Carriers**: is often used to denote individuals that are **heterozygous** for a particular **recessive** **allele**. They can have the allele, but do not express the **phenotype** associated with **homozygousity** of that allele. It can also be used to denote individuals that have the allele under discussion whether that allele is recessive or not.

**centiMorgan**: the unit of genetic **recombination** of the **chromosomes** is called a Morgan. A centiMorgan (cM) is 1/100th of a Morgan and is equivalent to 1% recombination. You can think of it as two **loci** that are 20 cM apart on the chromosome, have been measured to have a 20% chance of recombination between the loci during meiosis. In theory the maximum recombination distance that you can measure between two loci on the same chromosome is 50% (unlinked loci segregate at 50%) but due to the possibility of multiple recombination events occurring between two loci, loci even more distant that 50 cM can have their linkage detected because they will **cosegregate** more often than 50% of the time. A second recombination event occurring between two loci cancels the first one and the two distant loci are back in the same linkage that they had in the parent. You need markers that fall between the two distant loci in order to detect the double recombination events and determine the true linkage distance. For most human chromosomes one cMis equal to about a million base-pairs of **DNA** sequence. For chickens one cM is equal to around 300,000 base-pairs on the **macrochromosomes** and 25,000 to 50,000 base-pairs on the **microchromosomes**.

**Chromatin**: the **protein** and **DNA** complex that winds up the DNA in the **nucleus** of eukaryotic cells is called chromatin. <https://en.wikipedia.org/wiki/Chromatin>

**Chromosome**: humans have 23 pairs of chromosomes contained within the **nucleus** of **somatic** cells. Chromosomes contain the genetic material. Each chromosome is a double helix of **DNA** wrapped up in protein that allows the DNA to be coiled up in the nucleus. Chickens have 39 pairs of chromosomes. 29 are classed as microchromosomes (small chromosomes) with the 10 pairs of macrochromosomes accounting for over two thirds of the DNA content of the chicken genome. The 29 microchromosomes account for less than 1/3 of the chicken genome. <https://en.wikipedia.org/wiki/Chromosome>

**Chromosome**: are **protein** **DNA** structures found in the **nucleus** of a cell that contain the genetic material of the cell. Chromosomes were found to be inherited like **Mendelian factors** and the **chromosome theory of inheritance** was developed. Essentially, the Mendelian genetic factors reside in the chromosomes. DNA was found to be the genetic material, and chromosomes can be thought of as protein structures that wind up the linear DNA into compact units that we call chromosomes. Humans have 23 pairs of chromosomes (**diploid** organisms have two **genome** equivalents in their nucleus). There are 22 pairs of autosomes and a pair of **sex-chromosomes** for a total of 46 chromosomes in each human **somatic** cell.

**Chromosome theory of inheritance**: Chromosomes were observed to behave like Mendelian factors during meiosis, and that observation led to the theory that Mendelian factors resided on the chromosomes.

<https://en.wikipedia.org/wiki/Boveri%E2%80%93Sutton_chromosome_theory>

**Coding sequences**: are DNA sequences that code for **proteins** and are included in the **exon** sequences of eukaryotic coding **genes**. The intact coding sequences with the **introns** removed from the **transcript** are found in the mature **mRNAs** along with **untranslated** 5’ and 3’ sequences that flank the coding sequence of the mRNA.

**Codominance** is when the **heterozygote** has a **phenotype** intermediate to both **parental** **homozygotes**. The two **alleles** are considered to be fully **additive** in terms of the genetic contribution to the phenotype. For codominance all **genotypes** can be identified by their phenotype. The usual example of codominance for chickens is Andalusian Blue (the *Bl* locus). On a genetic background that would have produced a fully black feathered bird (*E*/*E* at the *E* locus) the *Bl* allele will dilute the black. *Bl*/*Bl* birds are mostly white feathered, *Bl*/*bl+* are gray feathered and *bl+*/*bl+* birds are black feathered. You can accurately predict what you will get when crossing birds within that population. If you wanted to produce all gray feather progeny you would cross a mostly white bird to a black feathered bird (*Bl*/*Bl* X *bl+*/*bl+* produces all *Bl*/*bl+*). If you cross mostly white feathered birds together you will get all mostly white feathered progeny. If you cross black feathered birds together you will get all black feathered progeny and if you cross gray feathered birds together you will segregate white : gray : black in the 1:2:1 genotypic ratio.

**Cosegregation**: **loci** **linked** on the same **chromosome** will **segregate** together more often than random dependent on how close they are to each other on the chromosome. Linked loci can only be separated and not segregate together by **recombination** between the two linked loci during **meiosis**.

**DNA**: deoxyribonucleic acid is the genetic material for most lifeforms that exist on earth. **RNA** is ribonucleic acid that has an hydroxyl group on the ribose sugar that deoxyribose does not have. DNA consists of polymers made of adenine, guanine, cytosine, and thymine deoxyribonucleotides that have a 5’ to 3’ orientation based on the sugar residue hydroxl group that is exposed on each end of the polymer. It exists as two complementary paired anti-parallel strands that form a double helix. Purines pair with pyrimidines, adenine pairs with thymine, and guanine pairs with cytosine (A-T, T-A, G-C, C-G). DNA is measured in base-pairs (bp) and the **transcribed** part of the gene can be over a million base-pairs in length while genes like *MC1R* have only around a kilobase-pair (kb) sequence transcribed. The double helix structure of DNA is used to accurately **replicate** the **chromosomes**. Each anti-parallel strand can be used to recreate the other strand. It means that **Mendelian** **factors** can be replicated along the chromosomes with fidelity.

<https://en.wikipedia.org/wiki/DNA>

**DNA methylation**: DNA can be modified by adding a methy group to existing bases after **replication**. Methylation seems to be associated with chromosome structure and transcription of the DNA, so it is a type of **gene** **regulation** mechanism. It is used in some bacteria to differentiate host DNA from pathogen DNA. (see **epigenetics**) <https://en.wikipedia.org/wiki/DNA_methylation>

**Dominant**: an **allele** is considered to be dominant if the **phenotype** associated with the dominant allele is expressed in the **heterozygote**, but the phenotype of the **recessive** allele is not. Mendel had his dominant factor and his recessive factor. The hybrid cross (*Y*/*y* X *Y*/*y*) for something like green and yellow seed color, produced a 3:1 dominant:recessive phenotypic ratio (3 yellow, *Y*/- and one green *y*/*y*) but a 1:2:1 (*Y*/*Y* : 2 *Y*/*y* : *y*/*y*) **genotypic** ratio. The dominant allele is designated with a capital letter. **Gene knockouts** are the easiest examples of dominance to understand. If a **mutation** destroys the function of a **gene** the normal functional allele is dominant over the gene knockout recessive mutation because **heterozygotes** have one functional gene, but it takes **homozygousity** of two gene knockout alleles to completely lose gene function in such individuals.

Both copies of a gene have to be knocked out to lose all gene function. This is why most organisms are **diploid**. There are two copies of the **genome** in every **somatic** cell. Each genome can have defects or altered function in some genes, but they can still have a normal copy of the gene in the other genome copy. This allows the existence of a vast quantity of **genetic** **variation** to exist within any species and has been the fuel for biological evolution.

There is **incomplete dominance** that can occur when one functional copy is not enough to maintain the normal phenotype, or in the case of **parts poisoning** when a defective gene product can interfere with the normal function of some cellular system, and can knock out the system when a single copy of a defective gene is present. If the system is incapacitated when the defective allele product is present, the defective allele is dominant, but if the system is only partially incapacitated you can get incomplete dominance.

**Epigenetics**: is the study of **heritable** traits, or a stable change of cell function, that happen without changes to the DNA sequence. The DNA sequence (order of the nucleotides along a strand of DNA) is referred to as the primary structure of DNA. The secondary structure would be the double helix, and the tertiary and quaternary structure is the protein-DNA complex that makes up the **chromatin**-**chromosomes**. <https://en.wikipedia.org/wiki/Epigenetics>

The DNA sequence or **chromatin** structure is modified without altering the DNA sequence. It is a means of regulating **gene** **expression**, and the modification is accumulative as the individual ages. **DNA methylation** has been used to calibrate the epigenetic clock that can be used to estimate an individuals biological age by methylation status. <https://en.wikipedia.org/wiki/Epigenetic_clock>

In terms of classical genetic interpretations **genome** modification without changing the sequence has most likely been explained as part of the influence of the environment that the genome has to exist in. It has been treated as part of the environmental component of inheritance.

**Epistasis**: has a specific meaning in **qualitative genetics**. It is when a certain **genotype** at one **locus** masks or hides the **phenotypic** expression of other loci. The example often given in the definition of qualitative genetics is white feathers. When you have the **genotype** to produce a white feathered bird you cannot determine what alleles may be present at all the other feather color and feather color pattern **loci**. If a bird had the genetics to produce the multicolored feather pattern of a Speckled Sussex, but it was recessive white, the bird would be white feathered, and you would not be able to tell that it had the Speckled Sussex feather pattern. So the genotype producing white feathers is epistatic over all the other feather color loci. For **quantitative genetics** epistasis is used more generally to refer to any type of allele interaction between genes that alters the usual phenotype associated with those **alleles**. Like qualitative genetics one allele might mask the expression of alleles of other genes, but that is only one type of **gene interaction**. Various alleles of two or more genes might combine to produce a new phenotype. Alleles might interact to have a positive or negative effect on a particular trait that would not occur if only one of the interacting alleles was present.

**Eugenics**: is best forgotten, but history could repeat itself. <https://en.wikipedia.org/wiki/Eugenics>

**Eukaryote**: there are two major lifeforms on earth **prokaryotes** and eukaryotes. There are viruses, but they usually are not considered to be free-living, but parasitize prokaryotes and eukaryotes in order to reproduce. Prokaryotes are considered to be single-celled organisms though they can form colony structures and biofilms. Prokaryotes are defined as lacking membrane bound organelles like the **nucleus**, **mitochondria**, and organelles without **DNA** like lysosomes and the golgi apparatus. Bacteria are prokaryotes. What we call plants and animals are multicellular eukaryotes whose cells contain membrane bound organelles.

**Exons and Introns**: it turns out that in **eukaryotes** most **coding genes** do not have one continuous **coding sequence** that can be **translated** into **protein** sequence. There are intervening sequences called introns that have been inserted into the **coding sequence**. The fragments of that are retained in the creation of the mature mRNA **transcript** are called exons. The introns are removed (exons are spliced together) after **transcription** to generate the final mRNA coding sequence. The sequence removed from the initial **RNA** transcript can greatly exceed the sequence retained. One of the largest genes transcribed into RNA is the *DMD* dystrophin gene that has 79 exons. The transcript has been found to be 2 million nucleotides in length, but after the introns are removed a 14,000 nucleotide mRNA remains. <https://www.ncbi.nlm.nih.gov/gene/1756>

***F***: is the coefficient of inbreeding. It is calculated using the pedigree. It is a calculation of the average **IBD** (identical by descent) portions of genome an inbred individual has. Full sibs inherit one genome equivalent from each parent. Due to segregation they share half of that genome equivalent from each parent so half of their genome is the same (¼ from one parent and ¼ from the other). Full sibs share half of their genetics. When you cross full sibs the chance that their offspring will inherit the same bits of genome from both parents is 0.5 x 0.5 = 0.25 = *F*. 25% of the inbred genome is expected to be IBD.

<https://en.wikipedia.org/wiki/Coefficient_of_inbreeding>

**F1**: is the first filial generation. The parental cross produces the first hybrid generation (F1). F2 is the second generation after the parental cross, and F3 is the third generation after the parental cross.

**Family selection**: is a type of selection that is used to remove low **heritability** **traits** from the population. These traits are often due to **recessive alleles** or due to gene interactions (certain alleles at two different **loci** needed to be present to express the **phenotype**). If two parents produce an offspring with the trait being selected against, none of the progeny from that mating or other matings with the same sire are selected as breeders for the next generation. The parents are **carriers** for the causative genetics, and 2/3 of their progeny that do not express the trait can be considered to be carriers. Half the progeny of the affected sire with other females are considered to be carriers. This results in the loss of this families’ genetics from the population, and the loss of the 1/3 and ½ progeny that were not carriers among full sibs and half sibs, respectively. There can be a serious loss of valuable genetics from the population if family selection is resorted to, so it is basically a last resort to clear a population of an unwanted trait.

**Gametes**: sexually reproducing **eukaryotes** have a gametophyte stage in their life cycle. In order for both parents to contribute genetics to the progeny there needs to be a reduction in the number of chromosomes accomplished through **meiosis**. Meiosis takes a normal **diploid** **2N** cell through a series of cell divisions to produce a **haploid 1N** gamete. There are eukaryotes that can produce both **gametophyte** haploid sperm and egg cells, but for animals like us, males produce the sperm cells and females produce the egg cells. The 1N sperm and egg cells combine to produce a 2N **zygote** of the next generation.

**Gametophyte**: the **haploid** **1N** stage of the life cycle of **eukaryotes** and this phase of the life cycle is called the gametophyte generation. For some eukaryotes the gametophyte generation can be extended and the haploid cells can themselves multiply as free living cells before fusing to reform the **diploid** organism. For humans the sperm and egg cells are the haploid gametes.

**Gene**: the concept of the gene has changed over time, and there is no current single simplistic definition that is accurate enough for anyone to be satisfied with. There are many types of genes or DNA sequences that might be considered to be part of a gene. The wiki puts forward multiple ways to think about what a gene is. The **Replication: Transcription: Translation** definition has a list of the various types of RNA products that are produced from DNA sequences found in the genome. Some of the classic **loci** **phenotypes** have been found to not be due to genes, but to reside in **regulatory elements** that regulate the expression of a **coding gene**. Some of these regulatory elements have been found to reside within the **transcribed** regions of genes neighboring the gene being differentially regulated. I was involved in mapping 5 toes (polydactyly) in chickens replicating some previous work. This article is expanding on the previous research, including ours, and is open access (<https://anatomypubs.onlinelibrary.wiley.com/doi/10.1002/dvdy.22623>). It turned out that the Sonic hedgehog (*Shh*) gene was being regulated by a DNA sequence called the ZRS region found in an **intron** of the neighboring *Lmbr1* gene. A single base substitution in the ZRS DNA sequence resulted in the dominant 5 toe phenotype found in breeds like the Silkie chicken.

<https://en.wikipedia.org/wiki/Gene>

**Gene expression**: gene usually refers to a **DNA** sequence that produces an **RNA** product the **Replication: Transcription: Translation** description lists many types of RNA products that are **transcribed** using the DNA template. These genes are usually regulated so that the amount of RNA product that is produced can be controlled as to when and in what cell type the gene is expressed. A gene is expressed when transcription is allowed to proceed. Gene expression is a general term that can be used when talking about when a gene is turned on or turned off (is being transcribed or not).

**Gene interaction**: we are all the result of the interaction of all the sequence of our **genomes** including the bits that we have identified to be genes. Every new mutation has to function within what is already working so the interaction of genes is a given. What we often designate as gene interactions are when we can see two or more genetic **loci** that have something to do with a particular **phenotype**. The straightforward example of gene interaction is given in the definition of **Hybrid** (pea comb and rose comb **alleles** combine to produce walnut comb). We also have the examples of **epistasis** and **dominance**.

**Gene knockout**: mutations that interfere with the function of a gene product to the point that it does not fulfill it’s normal function are called gene knockout mutations.

<https://en.wikipedia.org/wiki/Mutation>

**Gene regulation**: Though there are many concepts out there for what a gene is we are usually talking about a DNA sequence that is transcribed into an RNA product. There is a complex system in place that regulates how much RNA is made, and what happens to those RNA products after they are made. Gene regulation also extends to how the gene products are themselves regulated. **RNA transcripts** have to be processed and modified. Once the mature mRNA of coding genes is produced it’s use as a template to code for protein is regulated, and these RNAs have a rate of “turnover” that depends on how quickly they are recycled (depolymerized by scavenging enzymes that chew up the RNA). There is even post **translational** regulation in the form of further processing the protein products produced from the mRNAs. Some **proteins** like those involved in blood clotting are not processed to their active form until they are needed to stop blood loss.

Recessive white comes up several times among these Jargon descriptions. The causative mutation for recessive white was found to be the insertion of a retrovirus into intron 4 of the tyrosinase gene. It was found that the retroviral insertion interfered with splicing and that an incorrectly spliced mRNA was produced in the skin that did not produce functional tyrosinase. What is interesting is that the mis-splicing is tissue specific. Pigment is not produced in the epidermis, but can be produced in other tissues, so the retroviral insertion is correctly spliced out in tissues other than the epidermis such as the retina (recessive white is a black-eyed white unlike the pink-eyed white of albinos). White Silkie chickens are usually recessive white and can have pigmented dermal tissue, muscles, connective tissue and bones. The retroviral insertion regulates the **gene expression** of the tyrosinase gene in a tissue specific manner.

<https://pubmed.ncbi.nlm.nih.gov/16457736/>

<https://pubmed.ncbi.nlm.nih.gov/17878441/>

**Genetics**: The science of genetics is the study of inheritance.

**Genetic Code**: read the wiki to learn about what the genetic code is. My comment on the matter is that the genetic code really is not the genetic code of life. It is only part of the code of life. What we call the genetic code is only the 3 nucleotide vocabulary used to order the assembly of amino acids into a specific **protein** sequence using **mRNA** as a template to “read” the code using **tRNAs** and create the amino acid polymer on the ribosomes. Biologists and geneticists have known for decades that we are not just the sum of our **protein coding** **genes**. Humans share the vast majority of their protein coding genes with other mammals (we all have a basic gene set with a minor fraction of genes that may be novel to the various lineages). We can take a lot of the human genes and put them into mice and they work just fine even though the protein sequence may differ by 10% or more. We can even recreate human genetic diseases due to single amino acid substitutions in genes by engineering the same amino acid substitution in lab animals to recreate the same symptoms in those animals. We have identified some amino acid substitutions in human proteins that are associated with phenotypes that make us human, but when we look at our domestic livestock under selection we estimate that around 80% of the genetic variation that we are selecting for or against is likely **regulatory**. We basically have the same genes as a chimp with some coding variants that differentiate us, but most of the difference is in how the genes are regulated. We only had the tech to sequence short proteins, but the first 10 proteins sequenced for chimps and humans had identical protein sequences. We may vary by around 2% in our DNA sequence for most of our **genome** compared to a chimp, but the coding sequence has been preserved and mutations have been selected against because the sequence difference between chimp and human coding sequences is only 0.7%. It can be argued that most of what we are is due to gene regulation and not the genetic code. In the age of genome sequences this fact is often ignored. There exists another heritable “code” in the DNA that is used to regulate the genes, and we haven’t spent enough effort researching the matter so that it is as well understood as is the protein code. <https://en.wikipedia.org/wiki/Genetic_code>

**Genetic polymorphism**: is a variant in the **DNA** sequence that is **segregating** in the population. The original **mutation** event that created the genetic variant usually occurred many generations ago, and the genetic variant has been transmitted genetically since it’s origin.

**Genome**: usually refers to the **haploid** complete **chromosome** complement found in the **nucleus**. **Diploid** species such as humans have two sets of chromosomes in the cell nucleus. The complete diploid set of chromosomes is often referred to as the genome of an organism. For diploid species that have **sex-chromosomes** that contain different **gene** sets like the human X and Y chromosomes the full diploid set of chromosomes is technically the full genome even though human diploid cells are referred to as containing two genome equivalents. Human males technically have the full human genome in terms of gene content. Females with two X chromosomes lack the genes only found on the Y chromosome. Animals also have a cytoplasmic mitochondrial DNA genome found in the mitochondria of a cell. Plants have mitochondrial genomes and chloroplast genomes. Energy generation by oxidative phosphorylation is a function of the mitochondria, and it also does things like make amino acids and fatty acids. Photosynthesis occurs in plant chloroplasts. I earned my PhD sequencing the mitochondrial genomes of invertebrates. <https://en.wikipedia.org/wiki/Mitochondrion>

<https://en.wikipedia.org/wiki/Chloroplast>

**Genotyping array, SNP chips**: technology was developed that allowed massively parallel genotyping of individual single nucleotide polymorphisms (**SNP**). Hundreds of thousands of SNP could be **genotyped** at the same time that would cover the **genome** with **genetic variants** spaced every few thousand base-pairs. It allowed linkage determination between genes and SNP markers, and allowed mapping **phenotypes** to specific **loci** identified by their **linkage** to nearby SNP.

<https://en.wikipedia.org/wiki/SNP_array>

**GWAS**: **genome wide association study**: **PCR** allowed typing molecular markers such as **VNTR**s (variable number tandem repeats) and genotyping hundreds of markers spaced along a genome GWAS became more economical, though still pretty much cost and time prohibitive. GWAS was relatively rare until **SNP chips** or **SNP arrays** made it possible to type thousands of markers across the genome at one time. GWAS is dependent on linkage between the markers and the loci that affect the trait under study (see **linkage mapping**). Spacing markers along the chromosome is like making a fish net that can capture the targets that exist between the genetic markers along the chromosome.

<https://en.wikipedia.org/wiki/SNP_array>

<https://en.wikipedia.org/wiki/Genome-wide_association_study>

**Genomic Selection**: is an advanced form of **marker assisted selection** (**MAS**). The development of the “**SNP chips**” that could genotype thousands of genetic markers at the same time allowed the entire genome to be placed under surveillance for selection purposes. Instead of being able to track just a few markers used in MAS, genomic selection could select on the whole genome, at least, as much of the genome that was covered by markers. In theory genomic selection can be more accurate than **BLUP** because instead of relying on statistical estimates as to how related an animal is to their parents and more distant ancestors you can generate a genomic relationship matrix and determine just how related each individual in the population is to one another. BLUP just assumes that you inherit half of what each parent can contribute, ¼ from each grand parent, but it can’t tell you how much you actually inherited from those ancestors. Due to the randomness of **segregation** you can be more or less than ¼ related to each of your grandparents, but DNA is also inherited from ancestors shared by both parents, so that the same genetics can come from both parents (**inbreeding** and **identity by descent**) The genomic relationship estimate tells you what genetics you share with each individual in the population. It may sound odd, but you can be more than half genetically related to one parent because of what you inherited from your other parent. You can inherit genetics shared by both parents that they inherited from a common ancestor. The same shared ancestral genetics can be inherited from either parent. Both of your parents may have inherited the same portion of a chromosome from a shared great great gandparent. Your parents have a 50% chance of transmitting this bit of chromosome to their progeny. It may turn out that one of the progeny do not inherit the bit from their father, but do inherit the bit from the mother. For the genomic relationship estimate this chromosomal bit from the mother makes this progeny look more than 50% related to their father. It is an extra bit, the same sequence as some paternal DNA, but was inherited from the mother. <https://en.wikipedia.org/wiki/Genomic_selection>

**Genotype**: the genotype is essentially the primary sequence of DNA along the chromosomes. DNA sequence genotypes can be denoted with **allele** designations. *MC1R\*B/MC1R\*B* (*eb/eb*) is a genotype denoting the same allele on the two **homologous** **chromosomes**.

**Haploid, diploid, triploid, tetraploid** **etc**: ploidy refers to the number of **genome** copies a cell has. Normal diploid **eukaryotes** such as humans have 2 genome copies per cell. Diploid cells are usually denoted as 2N. The sperm and egg are haploid and have only one copy of the genome and are 1N. There are some organisms like the parthenogenetic lizards on some islands that are triploid (have 3 genome copies, 3N). These lizards are all female and do not need a male to reproduce. Tetraploid organisms have 4 genome copies in their cells (4N). Allotetraploids are often “instant” species. Two different species produce viable progeny even though they may not be very closely related and their **chromosomes** do not pair properly during **meiosis**. These hybrids are usually infertile, but if the cells become tetraploid they now have 2 copies of each parental genome (total 4N). Meiosis can proceed normally and they can produce viable **gametes**. The common ancestor of all vertebrates (includes humans) may have been an allotetraploid with four genome copies in each cell. This ancestor existed over half a billion years ago, and a lot of bits of each species genome have been lost over time. This ancestor had two complete diploid sets of chromosomes that may have come from two species, but extant vertebrates retain only remnants from both genomes, and are considered to be diploids except for lineages that have duplicated their genomes more recently. For plants that can **self-fertilize** it is instant speciation because any mating back to either diploid parent species results in triploids that have issues producing viable gametes. The most common cultivated wheat is an allohexaploid (3 different genomes doubled for a total of 6 genomes per cell). You can google “highest ploidy records” and you can get examples of plants with 1440 chromosomes.

**Haplotype**: A **chromosome** is one haplotype. Specific alleles of the loci linked on that chromosome exist as that specific haplotype for that chromosome in that individual. Loci can have local haplotypes associated with them because closely **linked polymorphisms** **segregate** together most of the time. Portions of a chromosome can have a specific haplotype. Long **IBD** segments of a chromosome are haplotypes inherited from a recent common ancestor. Short IBD segments of a chromosome are haplotypes inherited from a more distant common ancestor where there have been more intervening generations of **recombination**.

**Hardy-Weinberg equilibrium**: is used to estimate **genotype frequency** if the **allele** **frequency** is known for populations that are under random mating in the absence of selection. If those conditions are not met it can still produce a ballpark estimate for **genotype** frequencies, but can be pretty far off for populations under **selection** and under **assortative matings**. p2 + 2pq + q2 = 1 where p and q are the **major** and **minor allele** frequencies for a **locus** with 2 alleles. An example would be p = 0.9 and q = 0.1. The equation allows you to calculate the frequency of **homozygous** individuals for the major allele p2 = 0.92 = 0.81, the **heterozygote** 2pq = 2 x 0.9 x 0.1 = 0.18, and the minor allele homozygote q2 = 0.12 = 0.01. 81% are pp homozygotes, 18% are pq heterozygotes and 1% are qq homozygotes. You can use the equation to estimate the allele frequency in the population. If only 1 out 100 people in a population express a recessive trait q2 = 0.01 and the recessive allele frequency q would be the square root of 0.01 = 0.1. This would mean that there are 2pq **carriers** that are not expressing the trait in the population (2 x 0.9 x 0.1 = 0.18 or 18% carriers). If you see 1% incidence of a recessive trait that means that 18% of the population carry the recessive allele, but do not express the phenotype.

<https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle>

**Hemizygous**: a **gene** or whole **chromosome** having only one copy in a cell is considered to be hemizygous for that gene or chromosome. In the case of sex-chromosomes there is often a hemizygous sex. Human males have one X chromosome and one Y chromosome and are considered to be hemizygous for both X and Y chromosomes, but these sex chromosomes can still have genes that are common to both chromosomes, so they may or may not be hemizygous for specific genes on the sex chromosomes. The X chromosome is much larger than the Y chromosome, so most of the genes on the X chromosome are hemizygous in males (males have only one copy of the genes not found on the Y chromosome, but exist on the X chromosome). Human females have two X chromosomes, and so are not considered to be hemizygous even though females may lack genes that are only found on the Y chromosome. It is sort of strange to think about, but the human male with a normal X and the small Y chromosome is missing around 1.5% of the total diploid genome of a female, and yet it is the sex that has all of the genes that make us human (females are missing Y specific genes like *SRY*).

**Heritability** **in the narrow sense** (**h2**): To understand a lot of human genetic literature you have to understand the difference between h2 and **H2** (**heritability in the broad sense**). Heritability is the fraction of **phenotypic** variation that can be attributed to the genetics. H2 reflects the phenotypic variance that the whole genetic component **additive** and **non additive** combined is responsible for, and h2 reflects the phenotypic variance that the additive genetic component is responsible for. A lot of the older human genetic literature did not often make the distinction of what they were talking about, so their conclusions often missed the mark as to what they had found in studies like the inheritance of IQ where there is a huge environmental component (see the additive non-additive genetics description) and they were dealing with H2 and not h2. Animal breeders have had huge success using h2 estimates for selecting potential breeders because h2 is the most reliable measure of what genetics an animal will transmit to their progeny in order to make progress in the trait under selection.

**Heritability in the narrow sense** (**h2**) is an estimate of the additive genetic component (alleles that add or subtract from a trait when they are inherited). The higher the value of h2, the more similar the progeny are going to be to their parents.

**Heritablity in the broad sense**: (**H2**) is an estimate of the total genetic component of a trait both **additive** and **non-additive** genetics. A high H2 value for a trait does not necessarily mean that the h2 value will also be high. The non-additive genetic component is lowly heritable in terms of the phenotype inherited, and the trait may not be fully transmitted to the progeny in terms of being able to observe the genetics expressed in the **phenotype** of the progeny. The phenotype of the non-additive genetics that an individual has is not reliably transmitted to the progeny because the genetics are not additive. The non-additive genetics are inherited by the progeny, but they may not express as a visible phenotype. The additive genetics (h2) that an individual has is a better measure for how much the progeny will resemble the parent. The monozygotic twin (identical twins) studies all measure H2. h2 has to be measured by other means that are difficult to do in the human population. Basically, H2, because it includes the non-additive genetic component, does not tell you the value of the genetics an individual has to the population involved for the trait under study. H2 tells you what the total genetic component is for that trait for the individuals tested, but is not an accurate measure for what that individual can transmit to the population under test. Essentially, H2 is not a good measure for estimating what that individual can contribute to the next generation as an observable phenotype.

**Heterozygote and Homozygote**: A homozygote is an individual that inherited the same **allele** from both parents. It has two copies of the same allele (*MC1R\*E/MC1R\*E* or *E*/*E*). Heterozygous individuals have inherited different alleles from each parent (*MC1R\*E/MC1R\*B* or *E*/*eb*).

**Homologous chromosomes**: Your DNA is wound up into discrete packets that we call chromosomes. Humans have 23 pairs of chromosomes. 22 pairs of autosomes, and one pair of sex chromosomes. Human males have one X and one Y sex-chromosome and are considered to be **hemizygous** while females have two X chromosomes. Each pair are considered to be homologous chromosomes so humans have 22 pairs of homologous autosomes, and females have an additional pair of homologous X chromosomes. During **meiosis** the chromosome number is halved producing **haploid** **gametes** (sperm and egg) so that when the sperm and egg come together at fertilization the **diploid** number of 23 pairs is restored, and the resulting **zygote** inherits one of each pair from the two parents. The human genome is around 3 billion DNA base-pairs in length and there are two **genome equivalents** in each **diploid** cell (over 6 billion base-pairs). If you laid out both sets of chromosomes end to end you would end up with a little over 2 meters of DNA. All of this DNA is wound up in a **nucleus** with a diameter of around 10 micrometers creating 23 pairs of chromosomes.

**Hybrid**: is a term that needs to be taken in context of how it is used. Matings between two different species can sometimes produce hybrid progeny. The progeny of a cross between two distinct genetic lines are called hybrid progeny. You can have commercial hybrid products for crops and livestock that are the product of crossing two or more lines. Commercial layer hybrids are the product of crossing two sire grandparent lines to produce a hybrid male and two dam grandparent lines to produce a hybrid female. The hybrid male is then crossed to the hybrid female to produce a 4-way cross commercial product.

Hybrid can also be used to designate the production of **heterozygous** **genotypes**. The walnut comb example given in the definition of **qualitative genetics** is usually referred to as a **dihybrid** cross because if you cross a **homozygous** pea comb individual to a homozygous rose comb individual you produce progeny that are double heterozygotes for both pea comb and rose comb. PPrr X ppRR produces progeny that are all PpRr (walnut comb) dihybrids. If you cross these dihybrids together you generate genotypes depicted in many genetic textbooks using the **Punnett Square**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gametes** | **PR** | **Pr** | **pR** | **pr** |
| **PR** | **PPRR** | **PPRr** | **PpRR** | **PpRr** |
| **Pr** | **PPRr** | **PPrr** | **PpRr** | **Pprr** |
| **pR** | **PpRR** | **PpRr** | **ppRR** | **ppRr** |
| **pr** | **PpRr** | **Pprr** | **ppRr** | **pprr** |

Since both loci are **unlinked** they **segregate** independently. This can be easily observed using the pea comb **locus** to see that it segregates in the 1:2:1 (black(PP):blue(Pp):red(pp) letters) Mendelian genotypic ratio. The dihybrid cross **genotypic** ratio is 1:2:2:1:4:1:2:2:1 denoted by different color shading. Different shades of green are the singlets, different shades of purple are the doublets, and white is the quadruplet. The **phenotypic** ratio is 9:3:3:1 (9 walnut comb (P-R-), 3 pea comb (P-rr), 3 rose comb (ppR-) and 1 single comb (pprr)). <https://en.wikipedia.org/wiki/Punnett_square>

**Identity by descent**: **IBD** is due to the same allele being inherited from both parents due to the parents sharing the same ancestor. An individual is homozygous **inbred** at that locus. Due to the moderate rate of **recombination** IBD usually involves very large segments of a **chromosome**. It takes many generations to chop up the ancestral chromosome into small units of thousands of base-pairs that might encompass a single **locus** or **gene**. Around 60,000 years ago our modern human ancestors that migrated out of Africa interbred with a few Neanderthals so that non-african modern humans have a couple percent Neanderthal DNA our their genomes. Thousands of generations of recombination have occurred since the initial interbreeding and the bits of Neanderthal genome are now only around 25,000 base-pairs in length. We differ from Neanderthal by about 1 **SNP** (**single nucleotide polymorphism**) in 3,000 base-pairs, so these Neanderthal **haplotypes** can be identified by the more than half dozen Neanderthal specific SNP that they contain. Any individual that may happen to be IBD for a Neanderthal 25,000 base-pair **haplotype** is inbred for a very ancient ancestral sequence.

Recent paper on interbreeding of modern humans and Neanderthals: <https://www.science.org/doi/10.1126/science.adi1768>

**Imputation**: because **loci** are **linked** on the same **chromosome** and large **haplotypes** **segregate** in the population due to limited **recombination** during **meiosis** there are a finite number of haplotypes segregating within any population. These haplotypes can be identified by the **SNP** that they contain. It is not required that you identify all the SNP associated with a specific haplotype. All you need to do is identify enough SNP around the **genome** so that you can identify specific haplotypes. This means that if you can identify enough SNP to identify all the haplotypes in the population you can impute all the other variants associated with those haplotypes and so can impute the full genome from a subset of SNP **genotypes**. It usually takes enough sequencing to cover the genome around 30 times (30X coverage) in order to obtain the linear sequence for nearly all of the genome, but you can impute the genomic sequence using much lower sequence coverage of 1X or even less. It costs less than $300 to sequence the human genome to 30X coverage, but you can impute the genome sequence for 1/30 of the cost.

**Inbreeding**: is the mating of relatives. An inbred **locus** is defined as having **alleles** **identical by descent** (**IBD**). Meaning that the same allele was inherited from a common ancestor of both parents. Because there is limited **recombination** between **chromosome** pairs during **meiosis** there can be long sections of chromosomes IBD from matings between close relatives. Shorter sequences of IBD are due to inheritance from more distant ancestors. It takes many generations of recombination to reduce IBD sequences to less than a million base-pairs. ***F*** is the coefficient of inbreeding. *F* = 0 would mean the individual had no IBD loci, but *F* is usually estimated using pedigree rather than actual genetics so *F* is usually an under estimate of how inbred an individual is. In any population that has been closed for a reasonable number of generations inbreeding estimates based on recent pedigree misses the relationship to more distant ancestors. Today even a cross between two relatively unrelated commercial chicken lines producing **F1** **hybrids** would probably start with likely an *F* around 0.1 to 0.2 if you could genome sequence the F1 individuals and identify the IBD loci. This is due to the shared ancestors these commercial lines had before they were created as a subpopulation of domestic chickens, and subsequent inbreeding that occurred as separate populations. At the other extreme a mouse line that has been full sib inbred for over 25 generations likely has an estimated *F* > 0.99.

<https://en.wikipedia.org/wiki/Inbreeding>

**Inbreeding depression**: Inbreeding usually has a negative impact on the **phenotypes**. Inbreeding usually has a cost to the viability of the population that you are working with, so measures are taken to limit inbreeding in **pedigreed** populations. This is mainly due to deleterious **recessive alleles segregating** in the population. Mating relatives has a higher chance of creating **homozygous** recessive **genotypes**, exposing the recessive trait. See **Lethal load** for more explanation.

**Introgression**: an example of introgression is given in the description of **additive** **and** **non-additive** **genetics**. Crossing animals into a line to transfer a desired trait from the introduced animals is called introgression of that new trait when you have continued selection for that trait in the population undergoing introgression. **Backcrossing** the **hybrids** to other individuals of the line being improved transfers on average half of a **genome** equivalent from the animals first crossed into the line. The **F1** hybrids have inherited one genome equivalent from the donor animals, and their backcross progeny would inherit half of that on average. The first backcross progeny would have 75% of the recipient line’s genetics (they have 2 genome equivalents with half of one genome equivalent coming from the donor birds). If the donor trait is **dominant** or highly additive the first backcross progeny can be selected to have that trait, but if the donor trait is **recessive**, **progeny testing** is required to identify the carriers. This type of backcrossing can continue in order to recover as much of the recipient line’s genetics as possible and only transfer the donor trait to the recipient line.

**Lethal load**: or genetic load is the number of recessive lethal equivalents an individual member of a species or population may **carry** in their genome. For humans the average lethal load is estimated to be around 1.5 to 2. Obviously, everyone is not dead because we are **heterozygous** for the deleterious **loci**, and we are pretty much all carriers for some type of deleterious **phenotype**. This is why mating with relatives is a bad idea. Closely related individuals are more likely to have inherited the same deleterious **alleles** from their common ancestor. Labaratory Japanese Coturnix quail have been estimated to have a lethal load of 8.0, and no researchers have ever been able to create highly **inbred** (>0.99 inbred) research quail lines. Highly inbred chicken lines have been created, but not Japanese quail lines. All attempts have failed. All full sib mating lines started fail to produce a viable male and female full sibs within 4 inbred generations. Usually highly inbred lines are created by mating full sibs every generation, but these attempts have all failed with Japanese quail. A couple of efforts have gotten to around 0.5 inbred by cousin matings and selecting for viable matings, but my guess is that due to selection the inbreeding levels were not as high as estimated by pedigree (less inbred individuals were likely selected as breeders). So populations can look healthy and propagate very well with high lethal loads. You just can’t mate relatives without incurring **inbreeding depression**.

**Linkage**: the **chromosomes** contain a linear double stranded **DNA** molecule. All **loci** found on the same DNA molecule are linked. Linked loci will be inherited together unless **meiotic** **recombination** occurs between the two **homologous** chromosomes and new linkage is established. The recombinant will have loci that were once on the same chromosome transferred between homologous chromosomes creating two new linkage arrangements. Linkage is measured in units of recombination called Morgans. A **centiMorgan** (**cM**) is equal to 1% recombination distance.

**Linkage mapping**: in my opinion linkage mapping is one of the greatest developments in genetics. Linkage mapping and **DNA** sequencing is how we have identified the causative **mutations** and **genes** responsible for the classic genetic **phenotypes**. Linkage mapping is involved in the identification of the causative mutations of human genetic diseases. As a Post Doc I was involved in using linkage mapping to identify **QTL** (**quantitative trait loci**) for milk production. At the same time human genetics was using linkage mapping to identify genes like BRCA1 and BRCA2 that were associated with breast cancer. Sturtevant claims that he had the insight that creating a linkage order may have been possible, and he obtained the data and assembled the first linkage map over one weekend. Sturtevant was an undergraduate working in Morgan’s lab at the time. The phenotypes that he obtained data for did not **segregate** independently, they did not obey Mendel’s law of independent assortment. Sturtevant brought the data back in the form of the first linkage map of a **chromosome** (Sturtevant 1913). The genes were all on the same chromosome, but were located at different positions on that chromosome. Genetic **recombination** would break up this linkage at a frequency dependent on the distance that the genes were from each other on the chromosome. Genes closer to each other on the chromosome segregated together more often than those genes with a greater distance between them. The further apart two genes were the more recombination that could occur between them, so the genes could be ordered along the chromosome.

Any **polymorphism** that causes a measurable phenotype has a **locus** on a chromosome. Due to the phenomenon of genetic recombination the two **homologous** chromosomes can exchange genetic material, so a polymorphism that was found on one homologous chromosome can be transferred to the other, and that mutation now has a new linkage arrangement with the variation found on the other homologue. By determining the frequency of recombination between any identifiable points along the chromosome you can map the relative position of the locus you are interested in with the surrounding genetic variation. You can determine it’s position on the chromosome. All the alleles of the *E* locus map to the same location on the chromosome.

<https://en.wikipedia.org/wiki/Genetic_recombination>

**Locus**: is a location on a **chromosome**. There are multiple loci on a chromosome. **Linkage mapping** can identify the relative position of each locus on a chromosome.

**MAF**: is the minor **allele frequency** of a **locus** for 2 alleles **segregating** in a population. When two alleles of a locus are segregating in a population one allele is usually more abundant in the population and is designated the major allele because it is found at the highest allele frequency. The allele with the lower allele frequency is called the minor allele frequency. Why the minor allele gets the MAF abbreviation instead of the major allele is discussed in the **allele frequency** description.

**Marker assisted selection**: **MAS** is the use of genetic markers as tags to track the inheritance of specific alleles under selection. If we know the causative mutation for the difference between alleles we can usually create a molecular genetic test to identify that particular DNA variant. We can use these genetic markers to identify the progeny that inherit the allele under selection from a **heterozygous** parent. We can track recessive alleles and identify carriers in order to select for or against the recessive allele.

**Mendelian factors**: Mendel did not know anything about **genes** or **chromosomes**. What he described where things that he treated as discrete entities that were inherited as a single unit with fidelity. Even though you could not visually detect if the **recessive** factor had been inherited or not, if it was present in the individual the recessive factor could reappear unchanged among the **phenotypes** of the progeny of that **carrier** individual. Mendel’s factors were bits of something that could produce the phenotypes that Mendel was working with, and that these factors survived to **segregate** unchanged through the generations. Mendel didn’t know about **mutation**, and he never raised enough peas to identify new mutations. His factors were **polymorphic** loci segregating among sweet peas.

<https://en.wikipedia.org/wiki/Mendelian_inheritance>

**Meiosis**: is a process that produces the haploid sperm and egg cells with half the chromosome complement. I always wanted the students to understand meiosis because if you understand what happens to the chromosomes during meiosis you will understand why Mendel’s laws apply to genetics.

<https://en.wikipedia.org/wiki/Meiosis>

**Mitosis**: <https://en.wikipedia.org/wiki/Mitosis>

**Molecular genetic markers**: At the start of the rediscovery of Mendel’s work the genetic markers were the Mendelian **qualitative** trait **phenotypes**. When Sturtevant (1913) created the first **chromosome linkage map** he used the Drosophila traits that were then called “factors”. He used eye color, body color, and wing shape as tags to relate those **phenotypes** to their locations in the genome. Once **DNA** was identified as the genetic material any means of identifying specific DNA sequences could be used to create more direct genetic marker assays. When I was a graduate student we were initially limited to **restriction enzymes** and **Southern Blot** technologies. **PCR** was introduced and revolutionized genotyping. PCR allowed the typing of specific sequences and amplification of VNTR (variable number of tandem repeat) sequences. I was involved in the creation of the first **SNP** (single nucleotide polymorphism) chip high throughput chicken genotyping array (Groenen et al., 2011) and such genotyping arrays are now capable of genotyping hundreds of thousands of genetic markers at a time. The **next generation sequencing technology** has made it cheaper to perform low coverage genome sequencing in order to **impute a genome wide map** for each individual.

**Mutation**: is a heritiable change of the sequence of the genome. There are a lot of ways to change the genome sequence. Post replication base modification such as **methylation** of the DNA is not considered to be a mutation even though the base modification can be inherited for several generations. Base modification falls under **epigenetics**. One thing that I should say about mutation is that the term is often misused (even by myself) to describe a **segregating** **genetic polymorphisms** in the population. Genetic variants like the single nucleotide change in the *BRCA1* **gene** associated with breast cancer in humans is not a mutation, but it is a polymorphism or genetic variant segregating in the extant human population. The *BRCA1* type mutation may have occurred thousands of years ago, and has been passed down through the generations from the ancestor in which the mutation occurred.

<https://en.wikipedia.org/wiki/Mutation>

**Next generation DNA sequencing**: are the automated DNA sequencing methods that were invented to replace Sanger and Maxim and Gilbert DNA sequencing technologies.

<https://en.wikipedia.org/wiki/Sanger_sequencing>

<https://en.wikipedia.org/wiki/Maxam%E2%80%93Gilbert_sequencing>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC3841808/>

<https://en.wikipedia.org/wiki/Third-generation_sequencing>

**Normal distribution**: is a continuous probability distribution that produces the “bell curve” depicted in countless statistical figures. My story about distributions is as an undergraduate genetics majors at Berkeley my advisor thought that it would be good for me to take upper division statistics and probability theory from the math department. It nearly killed me, but I did end up learning what saddle points in calculus were good for. I did a couple years of undergraduate research in the genetics department and the professors that I asked all told me that genetics was over, and that molecular biology was the future, so I ended up doing molecular biology as a graduate student. By the time I got my PhD new **molecular genetic markers** had started to be used to map the genome and characterize quantitative traits. It turned out that genetics was not over, but was sort of starting over. So I decided to go back into genetics. In 1992 I got a post doc doing **QTL** mapping in dairy cattle and ended up on the first publication mapping QTL in a livestock species. Back then it was difficult to determine if we were doing what needed to be done. We were using the LOD score evaluations that medical geneticists were using at that time to map genetic disease traits, but when I asked what the LOD distribution looked like I was told that no one knew. Genotyping was labor intensive and the molecular genetic post docs had to do it along with the technicians. I was lucky enough to genotype the marker showing the largest effect. When I was told that it was a jackpot marker I went back to my genotypes and used old fashion graph paper to chart the progeny that had inherited one or the other alleles from the sire under test using X and O and plotted the quantitative phenotype expressed by those progeny in terms of what their daughters had inherited from them. Each allele had basically a normal distribution of phenotypes and the means were well separated. I became a believer. That paper (Georges et al., 1995) is the highest cited paper that I have participated in creating. That was research done 30 years ago and genetics is still not over even though we have had whole genome sequences to play with for nearly two decades. <https://en.wikipedia.org/wiki/Normal_distribution>

**Nucleus**: Eukaryotic organisms have cells that contain a membrane bound organelle called a nucleus. The nucleus of a cell contains the genetic material (**DNA**) and is where **RNA** **transcription** and post transcriptional modification of the RNA products occurs.

**Parental**: the parental **genotypes** are the genotypes of the individuals involved in the cross. A cross between two distinct lines is often called the parental cross. The parental cross produces the **F1** generation (first filial generation). Mating of F1 individuals results in production of the F2 generation. Often the F1 **hybrids** are mated back to one or both parental lines in order to observe **segregation** of **recessive** **alleles** in a 1:1 **phenotypic** and genotypic ratio. This type of cross is often referred to as a **backcross** and produces the first backcross generation.

**Parts poisoning**: is a form of **dominance** that occurs when a mutation produces a defective gene product that does not interact properly with the cellular system that it is supposed to contribute to. One bad part can mess up enough of the functional units to express the **mutant** **phenotype**. One of the best known examples of this is achondroplastic dwarfism in humans. A single amino acid substitution in the *FGFR3* **gene** messes up the receptors normal function. The receptor is a dimer made up of two *FGFR3* subunits (homodimeric) or one *FGFR3* subunit and another *FGFR* (such as *FGFR1* or *FGFR2*) gene subunit to produce a heterodimeric receptor. If only one *FGFR3* subunit is defective that kills receptor function, so enough receptors are knocked out to produce the dwarf phenotype (only ¼ of the *FGFR3* dimers would be functional). **Zygotes** that are **homozygous** for the *FGFR3* mutation die. So homozygous progeny are not observed in families with two dwarf parents. The reason that achondroplastic dwarfism is so common in the population is because 80% of the incidences are due to spontaneous **mutations** that occur in one of the parents **gametes**. 97% of those spontaneous mutations involve a single position in the *FGFR3* gene (amino acid residue Gly1138). Mutation at this location is estimated to occur in around 1 in 14,000 live births making that site having one of the highest mutation rates known. Surprisingly, there are other positions in the human genome with higher mutation rates. We observe the effects of new *FGFR3* mutation because it is dominant and has a large effect on the phenotype.

**PCR**: polymerase chain reaction is a technology that uses prior knowledge of the DNA sequence to amplify specific DNA sequences in any organism that uses nucleic acid as the genetic material. It can also amplify RNA by first using a reverse transcriptase step to produce DNA from the RNA sequence. <https://en.wikipedia.org/wiki/Polymerase_chain_reaction>

**Pedigree**: is a record of the family history of the individuals in a population. Most people know their pedigree back to at least their grandparents. Commercial breeders have maintained pedigree records for some of their lines for over 50 generations. Cobb started in 1916 and was selling pedigreed **Standard Bred** chicks. Back then breeders were selling chickens that had been bred to meet the physical standards of their breed. Cobb did well with his Barred Plymouth Rocks before the poultry industry started breeding for specific traits of commercial interest instead of breeding to a breed standard around the 1950’s.

<https://www.cobbgenetics.com/who-we-are#history>

**Phenotype**: is a physical attribute of an organism. It needs to be measurable so that variations in the phenotype can be classified as being associated with variant **alleles**. The major issue in genetics is that there is no such thing as one **gene** associated with one **phenotype**. Phenotypes are usually complex and are affected by many genes. Just think of how many genes affected Mendel’s tall and short pea plants. He had the luck to work with a major large effect gene for dwarfism, but many genes were involved in making the plants taller or shorter. I want to do another post on the complexity of the “simple” *E* locus phenotypes. Even when we get down to the study of single gene functions within the cell we find that gene expression is affected by many other genes active within that cell. The phenotypes that we observe are due to everything working together to produce that individual.

**Progeny testing**: is the genetic testing of an individual by producing progeny from that individual and evaluating the phenotypes of the progeny in order to determine the genetic merit of that individual. The example given in this series of descriptions is progeny testing for the recessive white allele (*c*) at the *C* locus. Individuals with colored feathers suspected of being **carriers** for recessive white can be crossed to a recessive white tester line. If the *c* allele is present in that individual half of the progeny will be white feathered and that individual under test would be determined to be a *C+*/*c* heterozygote.

**Protein**: is a polymer of amino acids also called a **polypeptide**. Proteins are produced by a process called translation where tRNAs are used to read the 3 base genetic code of the mRNA and produce the polypeptide on the ribosome. Proteins can be structural like collagen, enzymes like lysozyme, or signaling receptors like *MC1R*. <https://en.wikipedia.org/wiki/Protein>

**Punnett Square**: is a graphic depiction of the independent assortment of alleles to form genotypes in the next generation. A Punnett Square is used in the **Hybrid** description as part of the explanation of a dihybrid cross. <https://en.wikipedia.org/wiki/Punnett_square>

**QTL**: quantitative **trait** loci are locations on a chromosome associated with a quantitative trait such as human adult height. There are two major fields of interest in genetics, **qualitative genetics** (the study of inheritance of major phenotypic differences, Mendel’s unitary discrete units) and **quantitative genetics** (the study of inheritance of phenotypes that are affected by many loci scattered throughout the genome). If you can produce a large enough phenotyped population you can use **genetic markers** spaced around the genome to genotype the population and do linkage analysis between the markers and the phenotype. It is like crafting a fish net out of the genetic markers, casting it over the genome, and then seeing what you can catch in the net. You can only detect the loci with the largest effects if your phenotyped population is around 1,000 or so, but human genetics is now dealing with phenotyped populations numbering over a million. The largest effect locus may only account for 2 or 3% of the genetic variance associated with a trait with the vast majority of QTL having much smaller effects.

**Qualitative genetics**: deals with loci associated with large **phenotypic** differences between alleles. Qualitative genetics deals with discrete phenotypes associated with specific **genotypes**. These are the traits that Mendelian analytical tools can be used to characterize. Qualitative genetics deals with the alleles that can be considered to be dominant or recessive. There are also alleles at one locus that are **epistatic** over the expression of alleles at other loci. The classic example for chickens is recessive white (*c*) at the *C* locus. When a bird is *c*/*c* homozygous for recessive white it has all white feathers. Since all the feathers are white you cannot tell what *E* locus alleles the bird has. Recessive white is epistatic over all loci involved in altering feather color or feather color patterns. You cannot determine the genetics at those feather color loci on a recessive white genetic background. Qualitative genetics also can determine the interactions between qualitative genetic loci. The classic example of gene interaction in chickens is pea comb and rose comb. The dominant allele *P* is associated with the pea comb phenotype and the dominant allele *R* is associated with the rose comb phenotype. When an individual has both dominant alleles at the two loci a new phenotype appears that is called walnut comb. Individuals with the genotype *P*/*P* *R*/*R* will have walnut comb, but only one dominant allele is needed (a *P*/*p* would do just as well as the *P*/*P* genotype) so the dominant interaction of the walnut comb genotype is denoted as *P*/- *R*/- where the dash could be any allele. *p*/*p* *R*/- individuals have a rose comb, *P*/- *r*/*r* individuals have a pea comb, and *p*/*p* *r*/*r* individuals have the normal single comb phenotype. There is a **Punnett Square** representation of gametes and progeny genotypes in the definition of **hybrid**.

**Quantitative genetics**: deals with polygenic inheritance. Thousands of genes can contribute to the expression of a particular trait such as human adult height. The phenotype is quantitative in that there is a distribution of something like adult height within a population. This distribution can be a **normal distribution**

**Recessive**: an allele is recessive to the dominant allele if the phenotype of the dominant allele is expressed in the **heterozygote** excluding expression of the recessive phenotype. Essentially if the hybrid looks like one of the **parental** types, the missing phenotype is recessive.

**Recombinant inbred mapping lines**: <https://en.wikipedia.org/wiki/Recombinant_inbred_strain>

My comment about recombinant inbred mapping lines is that they tell us that inbreeding depression is a combination of **epistasis** and **dominance** issues, both have very significant effects, and that current methods of estimating epistasis and dominance inbreeding effects in selected lines are inadequate to identify the magnitude of such effects on animal breeding. All the inbred lines that go into recombinant inbred lines are highly inbred (over 99%) and they do not have reproductive issues. Some of the inbred lines have been selected to have larger litter sizes than wild populations. There are no recessive lethals left in these lines to impair reproduction, but when you put 12.5% of the genome from one inbred line into the genome of another many of the sublines each with a different 12.5% begin to fail upon inbreeding and trying to take the lines to homozygousity for the 12.5%. There has to be epistasis (interaction between genes) that is responsible for the decline in fitness, and the effects have to be significant because they need to resort to decreased inbreeding and selection to try to save these sublines and maintain an adequate coverage of the mapping genome.

Pretty much all the publications in animal breeding have claimed to find that epistasis and dominance, inbreeding effects are negligible in populations under selection and all of them have likely been wrong. In Muir et al., 2008 we did a genome wide inbreeding analysis for commercial and some standard bred lines. It is not mentioned in this paper because we did not provide the pedigree information, but the inbreeding estimates were much lower than they were expected to be based on decades of pedigree selection. In animal breeding the goal has been to maintain inbreeding levels to 1% or less per generation. After **BLUP** came in you had to adjust the selection of breeders so that acceptable levels of inbreeding were maintained because BLUP tended to select related individuals. The chicken breeds that the industry had started with were already inbred to a certain degree because that had been necessary in order to breed the lines to a physical standard (The American Standard of Perfection). Before the dawn of the modern broiler industry in the 1950’s if your birds did not meet the Standard your customers would think that something was wrong, and at the time nearly all the breeders were selling Standard bred stock. So the industry started with a certain inbreeding level, and the industry added around 1% per generation and still most of the broiler lines had inbreeding estimates of less than 0.4 in 2008. Somehow we had selected against inbreeding and had maintained a lot of genetic variation in our lines under very intense selection. My take is that phenotyping was accurate enough so that we could select for the least inbred birds from the best families. This would mean that dominance and epistatic effects of inbreeding could be differentiated between sibs by phenotype.

**Recombination**: **meiotic** recombination is the exchange of **DNA** between **homologous** **chromosomes** during the production of **gametes**. All **loci** on a chromosome are linked on the same DNA molecule, but meiotic recombination breaks up this **linkage** and allows exchange of DNA between the homologous chromosomes creating new linkage arrangements. The further apart two loci are from each other on the chromosome the more likely that recombination will break up the linkage and the loci will not be inherited together.

<https://en.wikipedia.org/wiki/Genetic_recombination>

**Replication: Transcription: Translation**: These three processes are considered to be the central dogma of molecular biology (**DNA** to **RNA** to **protein**). DNA is replicated by unzipping the double helix and using each single strand as a template to recreate the complementary strand. RNA is transcribed using the DNA as a template to produce mRNA (messenger RNA), rRNA (ribosomal RNA), tRNA (transfer RNA), lncRNA (long noncoding RNA), snRNA (small nuclear RNA), snoRNA (small nucleolar RNA), miRNA (micro RNA), piRNA (piwi-interacting RNA), and there are circular RNAs. Translation is the production of the protein product, and translation occurs outside of the nucleus on the ribosomes. The mRNA is used as a template and tRNAs are used to read the **genetic code** and create a polymer of amino acids in the correct order.

<https://en.wikipedia.org/wiki/DNA_replication>

<https://en.wikipedia.org/wiki/Transcription_(biology)>

<https://simple.wikipedia.org/wiki/Translation_(genetics)>

**Selection**: results in nonrandom mating. Without selection all individuals of a population would have an equal chance of producing progeny in the next generation. Natural selection is when some environmental factor makes it more likely that certain individuals will produce progeny. In animal breeding we perform artificial selection and pick the animals that we want to use as breeders.

**Self-fertilize**: Plants and some animals can produce both male and female gametes. Self-fertilization in plant genetics is often referred to as selfing. The pollen of a plant is used to fertilize the egg cells of the same plant. This is **inbreeding**. Instead of sharing half the genome as do full sibs a selfing plant shares all it’s genome, so half the genome becomes **identical by descent (IBD)** in one selfing (***F*** = 0.5). If selfing is continued you end up with highly inbred lines where they can be **homozygous** for just about all loci in the **genome**. Their **alleles** become identical by descent. Mendel had pure lines in the peas that he used because pea plants naturally self-fertilize, and you have to open the flower and transfer pollen from another plant in order to make crosses. So Mendel started with inbred plants that were homozygous and true breeding for the traits that he decided to use for his experiments.

**Segregation**: reflects the inheritance of **alleles** in accordance with Mendel’s law of independent assortment. Alleles **segregate** among the gametes (sperm and eggs). For an individual **heterozygous** for the pea comb **locus** (*P*/*p*) half the gametes inherit a *P* allele and half a *p* allele. Unless there is a mess up in meiosis both alleles are not usually inherited from the same parent. In effect the P and p alleles segregate from the heterozygous parent. For a double heterozygote for pea comb and rose comb (*P*/*p* *R*/*r*) you expect ¼ *P R*, ¼ *P r*, ¼ *p R* and ¼ *p r* among the gametes produced (see the Punnett Square in the **Hybrid** description). Segregation is a term that can be used to denote the fact that an allele can exist with other alleles at the same locus in a population. An allele is segregating in a population if not all members of the population have that allele. A mongrel backyard flock of chickens might be segregating for all of the *E* locus alleles.

**SNP**: single nucleotide **polymorphisms** are just that. They are **genetic variants** that involve a single nucleotide’s position on the **chromosome**. The **DNA** double helix can be hundreds of millions of base-pairs in length for some chromosomes. Any base-pair position along a chromosome can segregate a polymorphism in the population. In the example below, sometime in the distant past a C/G base-pair mutated into an A/T base-pair. A lot of the SNP that we put on the first chicken SNP array were likely due to mutations that had occurred in the ancestors of the progenitors of domestic chickens and were already segregating in the Red Junglefowl population before domestication.

5’AGTGCACTCGGAT3’ 5’AGTGCATTCGGAT3’

3’TCACGTGAGCCTA5’ 3’TCACGTAAGCCTA5’

**Somatic cells**: make up the organisms body. Somatic cells have a complete set of chromosome pairs. Somatic cells propagate by **mitotic** cell division. Somatic cells are often defined as cells not sperm or egg cells.

**Standard Bred**: in the 19th century standard breeds started to be developed in the United States. Breeds were established that conformed to a particular phenotype. Chickens were bred to meet a standard. Poultry breeders had to conform to these standards in order to sell their birds as that particular breed. There were many more individuals involved in farming at that time and many of them would raise a flock of chickens in order to sell the eggs and produce some cockerals for meat production, The farmers would buy their favorite breed and expect to get what they paid for. The American Poultry Association maintains these standards. I can’t recall how long I have been a member. I started seriously raising birds in my backyard as a graduate student but had raised birds (chickens, quail, and parakeets) since junior high school. I ended up doing independent study on chicken genetics using the University of Utah library. I did a Post Doc in chicken genomics, and ended up as an assistant professor in the Poultry Science department of the University of Arkansas before going into private industry as a molecular geneticist for Cobb-Vantress.

<https://amerpoultryassn.com/about/>

**Standing genetic variation**: is the genetic variation **segregating** in a population that may be useful for evolutionary studies and comparisons between populations in terms of their genetic diversity. For a population like humans where there are billions of individuals, you are dealing with the fact that every site in the human genome has likely been hit by a new mutation 30 to 60 times in just the extant population, so we do not count every possible sequencing error that we come across as part of the standing genetic variation. I’ve seen the cut off at **MAF** (minor allele frequency) less than 0.1% (1 out of 1,000). Since you have two alleles per person that means that the rare allele can be found in 2 out of 1,000 people. Most species comparisons do not genotype enough individuals to find many alleles with a frequency of 0.001. The standing variation is such that each of us differs from a relatively unrelated human by around one **SNP** (single nucleotide polymorphism) every 3,000 base-pairs. If you sequenced the genome of 50 relatively unrelated humans you might find a SNP around every 1,000 base-pairs. Humans are actually deficient in standing genetic variation. We have around 1/5 the genetic variation found in most other species. We seem to have suffered a population bottleneck, and either nearly went extinct or formed a new species out as few as one thousand individuals. As decimated as chimps are they still have around 3 times the standing genetic variation found among humans. Domestic chickens suffered a domestication bottleneck, but likely interbred with local junglefowl from time to time and got some of the genetic variation back. If you sequence the genomes of around 50 domestic chickens you might find a SNP around every 100 to 300 base-pairs depending on what breeds of chickens you include among the 50. When we created the Standard Breeds of chickens we greatly reduced the genetic variation in those populations, but we could get most of that variation back by intermating all the standard breeds. When you breed to a standard you obviously have to select for specific genetics and **inbreeding** is involved, and you can rapidly loose genetic variation by inbreeding, but the different breeds retained different sets of genetic variation, so even though each breed might have less variation than it started with you can recover the lost variation by crossing in other standard breeds.

<https://www.pnas.org/doi/epdf/10.1073/pnas.0806569105>

**Trait**: is just another term for **phenotype**. Trait has higher usage in **quantitative genetic** discussions than it has in **qualitative genetic** discussions.

**VNTR**: variable number tandem repeats are short stretches of DNA sequence composed of short tandem repeated sequences. Microsatellite VNTR are repeats usually 4 nucleotides or less. Single nucleotide repeat (AAAAA), dinucleotide repeat (CGCGCG), tri nucleotide repeat (CGACGACGA) etc. VNTR can have longer repeats, just as long as they can vary in repeat number. The length of the repeat used is limited by the technology used to detect them. <https://en.wikipedia.org/wiki/Variable_number_tandem_repeat>

**Wild-type**: In genetics **phenotypes** are compared to what is considered to be normal. For animal genetics involving subjects as diverse as flies and chickens the phenotype used as a standard to compare new **mutations** to was the one exhibited by the wild species of the research subjects. In the case of the chicken the phenotypes found in the Red Junglefowl became the “wild-type” phenotype and was given the **allele** designation of (+). For the new **gene** allele designations the characters that were not numbers or letters were dropped from usage, so instead of the + sign the designation for the wild-type allele has become *N* indicating the normal allele.

**Zygote**: the zygote is the product of the sperm fertilizing the egg cell. The zygote develops into the **2N** next generation progeny of the parental sperm and egg donors.

**References**:

Andersson, L., Bed’hom, B., Chuong, C.-M., Inaba, M., Okimoto, R., and Tixier-Boichard, M.. 2020. Chapter 3, The genetic basis for pigmentation phenotypes. In. Advances in poultry genetics and genomics. Edited by Aggrey, S.E., Zhou, H. Tixier-Boichard, M., and Rhoads, D.D.. Burleigh Dodds publishing.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A. Steele, M.R., Zhao, X.. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing.Genetics. 139: 907-920.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC1206390/>

Muir, W.M., Wong, G.K.S., Zhang, Y., Wang, J., Groenen, M.A.M., Crooijmans, R.P.M.A., Megens, H.-J., Zhang, H., Okimoto, R. Vereijken, A., Jungerius, A., Albers, G.A.A., Taylor Lawley, C., Delany, M.E., MacEachern, S., Cheng, H.H.. 2008. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds.PNAS. 105: 17312-17317.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC2582254/>

Smyth Jr., J.R.. 1990. Genetics of plumage, skin, and pigmentation in chickens. In Poultry Breeding and Genetics. Ed. Crawford, R.D.. Elsevier Science. Pages 109-167.

Sturtevant A.H. 1913. The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. Journal of Experimental Zoology. 14: 43-59.

Open access: <https://www.nature.com/scitable/content/The-linear-arrangement-of-six-sex-linked-16655/>

American Standard of Perfection. Published by the American Poultry Association. Various copyright dates.

<https://amerpoultryassn.com/>

My Google Scholar link: <https://scholar.google.com/citations?hl=en&user=QP5FNzgAAAAJ>