

CHAPTER I

The Basics

Introduction

Genetics is the science of biological inference based on regularities in patterns of inheritance. It is a fundamentally analytical enterprise. It has become clear, especially in the last 50 years, that many important facts about biology can be uncovered using genetic analysis that cannot be learned in any other way. Genetic analysis is the best window we have into the workings of evolution, and few modern biologists would quarrel with the famous dictum of Theodosius Dobzhansky, a renowned 20th-century evolutionist: “Nothing in biology makes sense except in the light of evolution.”¹ Genetic analysis depends on abstractions that have generated a specialized language with a private vocabulary. It is the essential concepts, language, and vocabulary of genetics that we explore in this book.

Patterns of inheritance can sometimes be observed in natural populations: For example, blue-eyed parents produce only blue-eyed children, whereas some brown-eyed parents can produce both blue-eyed and brown-eyed offspring. Some interesting inferences are possible from such observations in natural lineages, but they are quite limited compared to the inferences that can be drawn from quantitative analysis of the results of genetic experiments (see below) with well-chosen parents, brilliantly introduced in 1865 by Gregor Mendel using pea plants in his garden.²

Mendel’s innovations were many, but the most important merit explicit mention:

¹ Dobzhansky T. 1964. Biology, molecular and organismic. *Am Zool* 4: 443–452.

² Mendel G. 1866. Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865*, Abhandlungen, 3–47. Mendel read this at the February 8 and March 8, 1865, meetings of the Brünn Natural History Society. William Bateson translated the paper into English in 1901. The best source for the paper today is MendelWeb (<http://www.mendelweb.org>), which gives both the original and Bateson’s translation side by side.

2 CHAPTER 1

1. The choice of garden peas. Like many plants, pea plants are normally self-fertilizing. However, gardeners can arrange that they are fertilized instead by plants of the same species that have different properties.
2. The choice of properties to study. Mendel selected properties, such as pea color and shape, that remain invariant for generation after generation when pea plants self-fertilize (a property called “breeding true”).
3. The crossing of true-breeding plants. Mendel crossed plants that bred true for one trait with others that bred true for one or more different traits and then crossed the hybrid progeny with each other. This procedure resulted, at each generation, in progeny that differed from one another with respect to the inherited traits.
4. Counting carefully the numbers of the different types of progeny. Mendel was the first to recognize that the ratios of different types of properties to emerge from a cross might reflect the mechanism of inheritance. Accurate quantitation was thus of paramount importance. To this end, he produced and counted large numbers of pea progeny, the better to estimate the ratios of the different inherited properties.

Mendel’s quantitative experiments marked the dawn of our understanding of genetic principles and opened up lines of biological inference that ultimately illuminated virtually every aspect of biology. Following Mendel’s lead, today’s experimental geneticists organize the study of inheritance patterns, and thereby the association of DNA sequences with traits in genetic model organisms, by choosing suitable parents and subjecting them to crossing and breeding schemes that will yield the desired genetic information most efficiently. The remarkable success over the last century of such genetic analysis in plants, flies, fungi, bacteria and their viruses, worms, and mammals all depended on the ability to choose the parents and follow the inheritance of their traits across several generations of progeny.

Human geneticists cannot simply choose parents or arrange crossings that will be informative. Instead, human geneticists adopted methods that allow one to search through the population for families that might yield information about particular inherited diseases. These methods rely on DNA technologies developed only recently, in the last two decades of the 20th century. Today, genetic inference is being increasingly used to discover and analyze inherited human disease-susceptibility genes by connecting them to differences in DNA sequence. We now can look forward to a day, in the relatively near future, when a routine medical examination will begin with the patient’s genomic DNA sequence, which, as we shall see, is only interpretable through the lens of genetic analysis.

Whether one is analyzing arranged crosses in experimental animals or existing human families, an important feature of study design is an estimate of how much data will be required to achieve statistical significance, because the way in which DNA is inherited is fundamentally probabilistic. Mendel intuitively chose the number of progeny to count; today, sophisticated statistical estimates of power using computational methods are standard. It is worth noting that many of the theoretical advances in the fields of probability and statistics since Mendel's time were motivated by the continuing desire to make genetic inference more rigorously quantitative.

Extracting biological insight from crossing and breeding studies depends on our understanding and applying a small number of abstract ideas. Geneticists developed special words to represent these ideas, which, as we shall see, are not always readily conveyed in ordinary language. The invention of new words (or the appropriation of existing words) into a technical language that can express genetic abstractions began with Mendel himself and has continued ever since. In some ways, genetic concepts resemble those of physics and chemistry, although the diversity and range of biological individuality, as well as the role of chance, make the extraction of mathematical regularities from biological phenomena particularly challenging. In the end, mathematical formulae often turn out to be less useful in dealing with genetic abstractions than are well-defined words of special meaning, combined with a few stylized diagrams.

The abstractions and concepts that are the most important and useful to genetic analysis are independent of any organism. This is not to say that all organisms deal with their DNA in exactly the same way. On the contrary, they are very diverse, differing in such basic things as the number of copies of their genome their cells normally contain. The analytical ideas emphasized here apply to every organism, even though not every kind of experiment is possible in all of them. Specifically, some organisms (viruses, bacteria, plants, and some fungi) are easily manipulated in huge numbers, whereas others (whales, elephants, and, of course, humans) are not experimental organisms at all. Nevertheless, the basic analytical ideas of genetics apply to each.

In writing this book, I have placed a great deal of emphasis on “implicit experiments”—experiments that underlie the definitions of the most central abstractions of genetics. Each of these is based on real experiments that I have idealized for the purpose of exposition. Real experiments depend on details and realities, many of which are specific to the organism under study. I have deliberately left out such details, particularly those I have found to be unhelpful and even distracting for the general reader seeking to understand the general principles.

For similar reasons, I have tried to avoid illustrative examples that require substantial background knowledge in the biology of specific organisms. Readers will no doubt notice that I favor examples from human genetics. I did this for two reasons. First, most readers will know much more about human biology than they will about the biology of bacteria, flies, or nematode worms. Second, most of what we have learned in other organisms applies also to humans; often the basic genetics was first worked out in one or more experimental model organisms. The success of this approach to understanding human biology is, quite properly, the well-justified basis for the continued societal support of research with model organisms.

Genetic Variation

All of genetics is based on studying variation in the genomic DNA sequences of individuals in a population. Individuals from every species naturally exhibit this variation; without it there could be no evolution. The level of genetic variation that exists in a natural population is a product of many factors. Notable among these are the rate of errors in DNA replication and the size and evolutionary history of the population. Some, but not all, of this DNA sequence variation has biological consequences and thereby contributes to the biological individuality of each member of a population. Of course, there are other equally profound contributions to individuality that arise from the environment and from the differences in the history of each individual.

Genetic inference is based on the outcomes of cross-breeding individuals of the same species, who differ both in their genomic DNA sequence and in one or more biological traits. Genetic analysis seeks to distinguish whether any differences in their DNA sequence influence one or more of these traits. Ultimately, for simply inherited traits (those that are attributable to the function of a single gene), this comes down to a conceptually simple question: Is a particular sequence variant inherited by progeny from one of its two parents always inherited with the particular trait that is also inherited from that parent? When a statistically secure connection is found between a sequence variant and a trait, we infer that this genetic difference causes, at least in part, that particular biological trait.

Experimental geneticists today no longer depend on natural variation for determining whether a DNA sequence is linked to a trait. This is because biologists have developed increasingly efficient methods of artificially inducing random alterations in DNA sequences for use in experimental organisms since the late 1920s. More recently, advances in DNA technology have made it possible for researchers to construct experimental organisms with specific DNA sequence

changes at will. These advances have simplified the process of connecting DNA sequence differences with their biological consequences in these organisms. However, even these relatively straightforward experiments ultimately require the abstractions of genetic analysis for their interpretation.

Structure and Function

What can one actually learn from studying patterns of inheritance? The great inferences began with Mendel in 1865. Although we recognize his historical importance, it is hard, after so many years, to appreciate the true nature of Mendel's contribution today, when faced with the huge body of knowledge about inheritance that has been discovered since his day. Mendel knew nothing about DNA, chromosomes, proteins, or even “information” in the modern sense (i.e., something that can be reduced to and faithfully transmitted as a string of binary digits). Nevertheless, it is possible to restate his central insight in modern language. Mendel realized that the experimentally reproducible patterns of inheritance of well-chosen traits in pea plants meant that each plant contains two copies of the DNA sequences that cause each of these traits, one inherited from each parent. He also realized that each plant passes one copy of that DNA information to each of its progeny. Each of the seven traits Mendel chose to describe in his paper behaved independently of the others in experiments involving more than one trait at a time.

Mendel called the causative DNA sequences “factors” (he used the German word *Anlage*, which has additional connotations in English, including “predisposition” and “potentiality”). There are two classes of questions one can ask about Mendel's factors that are of enduring interest. One concerns the *structure* of the factors. Today, we know that these factors are stretches of DNA sequence that reside at particular positions on chromosomes and that they can be tracked on the basis of their positions. The second concerns the *function* of the factors. Today, we know that the information encoded in these DNA stretches can affect the biology of an organism in many ways—most often by specifying a protein or RNA molecule that actually fulfills particular functions at prespecified times and places in the life of an organism.

The pea plant (*Pisum sativum*) that Mendel worked with is, in today's language, “diploid,” meaning that the cells of the mature organism contain two copies of every chromosome (except for sex chromosomes) and thus two copies (not necessarily completely identical) of every genomic DNA sequence. As indicated above, the central inference of Mendel's paper is that during sexual reproduction,

the reproductive cells of the plant (its pollen and ovules, also called the gametes) contain only one copy of every genomic DNA sequence; in modern terms they are “haploid gametes.” This system of alternation between haploid and diploid is characteristic of most, if not quite all, eukaryotic organisms (including plants, fungi, humans, and animals). However, not all organisms are stably diploid; bacteria and viruses, notably, are normally haploid. Nevertheless, the basic concepts of genetic analysis apply to them all.

Gene and Locus

Distinguishing between structural and functional issues is fundamental to thinking clearly about genetics. Some of the confusion between the two different aspects of Mendel’s factors proved difficult to resolve. Thus, geneticists argued about the “nature of the gene” for decades. Today, most geneticists use “gene” and “locus” in ways that help to minimize this confusion.

Gene: In 1905, a Danish botanist and geneticist called Wilhelm Johannsen introduced the word “gene,” defining it in a way that subsumed both the structural and functional aspects of Mendel’s factors. Still today, the word “gene” is used quite generally in both a structural and a functional context. However, most of our thinking about genes today refers to function; the modern consensus about the functional interpretation of the gene dates back to the work of American geneticist and pioneering molecular biologist Seymour Benzer in the 1950s (whose work I discuss in more detail in Chapter 6). Thus, the use of the word “gene” to refer to sequence or positional information, which remains part of its meaning, has to be done with care. I will have much more to say about the functional definition of the gene in later chapters.

Locus: Many geneticists (including me) like to refer to a stretch of DNA as a “locus” when discussing its position in the genome, even when we know it’s a named gene with a known function. The word “locus” today unambiguously refers to position, and not to function. It is often used to refer to groups of genes that reside in the same region of a chromosome but whose functions may or may not be related to each other.

Genotype and Phenotype

The potential for confusion between functional and positional/structural considerations means it is important to keep separate in one’s mind variation in DNA

sequence from variation in biology among individuals. Sometimes, there is a causal relationship between the two; however, often there is no such causal relationship or it is a complex one. The earliest geneticists formalized this distinction (even though they knew nothing about DNA) by adopting two special words, “genotype” and “phenotype,” that serve to keep separate DNA sequence variation and its biological consequences. In the following section, I explain the modern working definitions of these words.

Genotype: This word refers to the DNA sequence of an individual. Different individuals have different genotypes whenever they differ in their DNA sequence, whether or not this difference has any known biological consequence. Genotype includes, in principle, the entire genomic DNA sequence of an individual, but is often used in a way that refers only to a particular gene or locus or a limited subset of genes and loci. A quick way to remember this is that in the context of genetic analysis:

- ▶ *Genotypes are differences in DNA sequence that distinguish individuals from one another.*

Phenotype: This word refers to any and all differing biological properties of individuals, but not including differences in their genomic DNA sequences. Today, the word is used very broadly to include any traits of an individual. In genetic analysis, phenotype tends to be used more specifically to refer to traits when they might possibly be the biological consequences of specific differences in genotype between individuals. In this usage, phenotype, like genotype, is often used in a way that refers only to a particular trait or subset of traits. A quick way to remember this is that in the context of genetic analysis:

- ▶ *Phenotypes are the visible or measurable properties/traits that distinguish individuals from one another.*

The introduction of these words (by the same Wilhelm Johannsen who had invented the word “gene” a few years earlier) was important because, like Mendel, other early geneticists could only infer genotypes on the basis of the inheritance patterns of phenotypes. These words therefore made it easier to avoid confusion, and their use has persisted even with the emergence of our current understanding of genes as DNA sequences. It is largely because of these words that the distinction between genotype (what is encoded in the DNA) and phenotype (the visible or measurable potential consequences of a genotype) has remained the same since the days of Mendel, even though much has been learned (and not just about DNA) since.

Enough specialized vocabulary has now been introduced to allow a rigorous definition of what genetic analysis is all about:

- ▶ *Genetic analysis relates genotypes to their phenotypic consequences and vice versa.*

Homozygote and Heterozygote

All crossing and breeding schemes involve the formation of cells that contain both of two alternative, parental DNA sequences of, for example, a locus or gene. In diploid organisms that reproduce sexually, every cell in the organism (except the sperm and the eggs, which are referred to as the “gametes”) contains two entire genomes, one derived from each parent. In bacteria, which are normally haploid, cells sometimes contain two different copies of only part of their genome. Experimental geneticists, beginning with Mendel, have deliberately produced hybrid organisms, which differ at particular loci or at genes of interest. In 1902, immediately after the rediscovery of Mendel, an English geneticist named William Bateson introduced two special words, “homozygote” and “heterozygote,” to distinguish between instances in which the two copies of DNA information in a cell are identical or different. These words are so useful that their usage quickly became standard and remains so today.

Homozygote: This word is used when the two alternative copies of genomic sequence present in a cell are identical. Geneticists often refer to homozygosity with respect to one or a few genes (or loci) and not necessarily with respect to the entire genome. Genetic variation is sufficiently common that the deliberate construction of an organism with two absolutely identical genomic sequences is challenging, even in model experimental organisms. Conversely, it is common, even in natural populations, for considerably long stretches of DNA (encompassing sometimes dozens of adjacent genes or loci) to be homozygous.

Heterozygote: This word is used when the two alternative copies of genomic DNA sequence are different. A cell or organism is said to be heterozygous whenever such differences exist. These differences can involve as little as one DNA base pair (bp) or as much as an entire chromosome. An important example of the latter is sex chromosomes. In humans, females have two complete copies of the X chromosome, but males have only one copy and contain a Y chromosome instead of the second X. Thus, only females can be homozygous at loci on the X chromosome. You should bear in mind that the two X chromosomes in females drawn from natural populations are generally not entirely identical because of the ubiquitous population-level variation in DNA sequence. Indeed, a substantial

level of heterozygosity throughout the genome is the hallmark of natural populations, which are continually evolving.

Hemizygote: This is a more recent term that refers to a subset of heterozygotes—those entirely missing a particular stretch of genomic DNA, as opposed to having a second copy that contains a divergent sequence. Human males are properly referred to as being hemizygous for sequences on the X chromosome. The American geneticist Hermann J. Muller in 1935 extended the common usage of this word to instances in which a single gene or locus is completely deleted in one copy of a diploid genome, and this usage has become quite common. The word is not used in cases in which only a part of a gene is missing; in such cases, “heterozygote” is the correct word to use.

DNA Variants: Mutations, Polymorphisms, and Alleles

DNA sequence differences have, over the years, acquired many diverse and sometimes confusing names. Before considering these, it is important to reemphasize that all DNA variants—whether they are natural or artificially induced, whether they affect only one DNA base pair or extend over many, whether they are common or rare—in the end are all just differences in DNA sequence.

Mutation: This word was introduced in 1901 by Hugo de Vries, a Dutch botanist and an early geneticist, to describe newly arising heritable changes in otherwise true-breeding varieties of plants. de Vries, one of the rediscoverers of Mendelism in 1900, recognized what he called mutant plants by noting new traits that turned out to be heritable in the ways described by Mendel. He inferred that there must have been a change in an underlying Mendelian factor. In modern language, we would describe this as inference of a change in genotype on the basis of an observed heritable change in phenotype. It is this change in genotype that de Vries called a mutation, and this has been the usage ever since.

Today, if we were to find such a mutation, and one that acts as a “dominant” mutation (see below), we would verify the above inference by sequencing the DNA of the mutant and comparing it to the DNA sequence of the (nonmutant) “wild-type” parent line. If, as we cross and breed this new mutant, we observe that whenever the new phenotype is found the variant sequence is also found, and vice versa, we have strong evidence that we have found the mutation that causes that phenotype. The breeding scheme is required because there is a considerable likelihood that sequencing the mutant would reveal the presence of more than one sequence change in the genome; where this is the case, it is the mutation that faithfully follows the phenotype in crosses that is the causative

one. When modern oncologists sequence human tumors in search of causative mutations, this is exactly the problem they face, except that the crossing and breeding approach is not available to them.

In modern usage, the word “mutation” can refer to virtually any change in DNA sequence. Nevertheless, the word is most commonly used to refer to rare alterations in a DNA sequence that cause a phenotypic consequence in organisms, one that makes them distinguishably different from the organism as it occurs in the wild (hence the term “wild type”). Today, the word “mutation” by itself always refers to genotype; if one wants to refer to a phenotype associated with a mutation, the correct and unambiguous usage is “mutant phenotype.”

For DNA sequence differences that derive from variation in natural populations, and especially those for which no phenotypic consequence has been found, I prefer the words “DNA polymorphism” or simply “DNA sequence variant.” The use of these words offers some advantages. First, unlike “mutation,” they are explicitly agnostic with respect to any potential phenotypic effect. Second, when they originate in natural populations, their characteristic phenotypes, if they have any, are actually all “wild type” (e.g., the different colors of flowers one might find in a field). In human genetics, we try to avoid referring to patients as “mutants,” even when it is fully justified scientifically; the word carries unfortunate cultural connotations. For example, we now know that hemophilia A is caused in males by mutations on the X chromosome that affected individuals inherit from their mothers. However, the locus that determines the synthesis of the clotting protein that is affected or missing in hemophilia A is highly mutable, and about one-third of all hemophilic males contain a causative DNA change that is not found in the mother’s DNA; these changes are new (rather than inherited) mutations. Nevertheless, we do not call hemophilic boys “mutants.”

Polymorphism: This is a word with deep roots in biology. The *Oxford English Dictionary (OED)* found a use of this word by the British naturalist Charles Darwin in 1846, only 6 years after its first use in any publication, a reference to diversity among portraits of the queen (so not used in a truly biological context). The plain English definition in the *OED* is “. . .the occurrence of something in several different forms.” This definition obviously applies straightforwardly to phenotypic variation in natural populations. Over the years, the word came to be used to describe variation in genotypes as well. Although the ambiguity between genotypic and phenotypic polymorphism is both unavoidable and acceptable up to a point, today it is usual to modify the word to indicate whether genotype or phenotype is meant. Of course, when the connection between genotype and phenotype is already well established, as, for example, in the case of blood group

antigens in humans, then there is no problem: Both genotype and phenotype are polymorphic, and current usage reflects this.

DNA Polymorphism: This usage of “polymorphism” is unambiguous and refers specifically to differences in DNA sequence. As indicated above, it is useful as an alternative to “mutation” when referring to natural populations or to humans. Originally, some authors had required that a particular alternative sequence was present at a minimum frequency in a natural population for it to qualify as being a DNA polymorphism. I think these academic arguments are best left behind. Today, I use “DNA polymorphism” whenever I need to refer to specific DNA sequence differences in populations but do not want to use “mutation,” “DNA sequence variant,” or “DNA sequence change.”

A method of mapping human disease loci using DNA technology, called linkage mapping (which I discuss further in Chapter 3), was first introduced in 1980. It takes advantage of the millions of DNA polymorphisms that are found at a relatively high frequency in human populations, most of which cause no phenotype. This abundance of functionally silent DNA polymorphisms also provides the basis for the various forensic methods that are used to identify individuals from samples of tissue or blood.

Several technologies that detect DNA polymorphisms came into general use during the 1980s as markers for genetic mapping. The DNA polymorphisms that each method detects were named differently, resulting in a zoo of acronyms. I do not think the differences between them are fundamental, but because these acronyms are used abundantly in the literature, I explain some of them here.

RFLPs (Restriction Fragment Length Polymorphisms): This approach detects sequence variants by using DNA-cutting bacterial proteins called restriction endonucleases, which recognize very short (4- to 10-bp) sequences of DNA, to cleave DNA into fragments of different lengths. These fragments can then be visualized as bands on gels that separate them by size—hence the name.

VNTR or STR (Variable Number Tandem Repeat or Short Tandem Repeat) Polymorphisms: This approach also uses restriction endonucleases to generate different-sized DNA fragments from polymorphic loci that feature tandem repetitions of short stretches of DNA: The variation is in the number of repeats. The RFLPs generated from such alleles can be particularly informative because these DNA loci often vary considerably between individuals, and many different lengths can be found in the population. As a result, markers for these polymorphic regions are used today in most forensic applications, notably in CODIS, the U.S. Federal Bureau of Investigation’s Combined DNA Index System, which

is used to analyze DNA profiles that can distinguish any individual human from all others except identical twins.

AFLPs (Amplified Fragment Length Polymorphisms): These polymorphisms are detected when cleaved fragments of DNA are amplified using polymerase chain reaction (PCR). This is a sensitive method that is also used to detect DNA polymorphisms for forensic purposes.

SNPs (Single-Nucleotide Polymorphisms): These polymorphisms consist of a single-nucleotide change and are detected using a variety of methods, most simply via the direct sequencing of DNA.

Genotyping: This word refers to the process of detecting DNA polymorphisms, regardless of the technique used. Essentially all DNA sequence differences can serve as genetic markers, and all classes of DNA polymorphisms can now be detected by direct DNA sequencing. This method is rapidly gaining favor given the precipitous reductions in the costs of new sequencing technologies.

Allele: This very useful word refers to mutants or variants in the DNA sequence at a single gene or locus. Like many of the important invented genetic words, its meaning has evolved over time as geneticists have wrestled to understand the nature of the gene. In 1902, William Bateson introduced the word “allelomorph” (shortened to “allele” in the 1920s) to describe the alternative heritable determinants (Mendel’s factors) that segregate away from each other in heterozygotes to produce the Mendelian ratios observed in their offspring.

During much of the early history of genetics (i.e., until the 1950s), the relationship between “locus” and “gene” remained murky. Throughout this period, evidence for the existence of alternative DNA sequences derived from the same locus or gene (also called allelism) consisted of a confusing mixture of functional tests and genetic mapping. The classic literature must therefore be read with considerable care around this point. Much (but not all, unfortunately) of the confusion was dissipated by the work of Seymour Benzer, which will be discussed in more detail in Chapter 6. Earlier literature regularly assumes that all mutations or variants that map to the same locus control the same function and, conversely, that mutations and variants that control the same function will map to the same locus.

Today, geneticists use “allele” to indicate any of the many alternative DNA sequences that can occur at a single genetic locus (that is, a single limited stretch of DNA sequence in the genome). Because the definition of “allele” relates to DNA sequence, alleles are, therefore, features of genotype. However, experimental geneticists who work with model organisms often define “allelism” by using a

functional test, called complementation (see Chapter 2), to show that putative alleles are alternative versions of the same functional gene. This inference assumes (with considerable justification) that all noncomplementing mutations that define a functional gene most likely reside at the same locus. In the discussion of complementation in Chapter 2, the reader will find that there are some rare exceptions to this generalization. Sometimes, direct DNA sequence evidence is used to associate differences in DNA sequence with a functional gene by showing that within these differences sit the known boundaries of a gene's sequence. This alternative form of evidence is particularly useful if the allele has no useful phenotype and thus cannot be tested functionally.

However, the DNA polymorphisms that can be used as genetic markers, especially in human genetics, generally have no phenotype and often do not sit within a functional gene. For this reason, allelism among such DNA markers is strictly a matter of differences in sequence (variants) along the same short stretch of DNA. Notably, some polymorphic DNA marker loci (the VNTR/STR subsets of RFLPs) are useful precisely because one finds multiple alleles at the same locus that are readily distinguishable from each other. For example, if such a locus contains a dozen repeats of a simple sequence, one can find up to a dozen distinguishable alleles for one locus, each containing a different number of repeats. This is what makes such polymorphic loci much more informative than SNPs in family studies or in forensic applications.

Functional genes, in any organism, can and frequently do display DNA polymorphisms in their sequences. Thus, several "wild-type" alleles may exist in nature, which will mostly have no known phenotype. Experimental geneticists use an informal convention whereby the unaltered, naturally occurring sequence of a gene or locus that is present in an organism used in genetic experiments is referred to as "wild type" to distinguish it from mutant alleles of the same gene. Readers might have already read in published genetic studies, for example, that a particular genetic mutation caused a phenotype that differed from the wild type, even though in nature the DNA sequence of the gene involved is polymorphic.

Much of experimental genetics concerns the manipulation and analysis of mutant alleles of functional genes. Sometimes, the mutant alleles are spontaneous or randomly induced mutations ascertained by their phenotypes. In more recent years, mutants have tended to be the result of the deliberate creation of DNA changes (often deletions) by genetic researchers. Mutant alleles of some genes can vary widely in phenotype, whereas others tend to have very similar phenotypes, differing only by small degrees. Rigorous genetic experiments require that any analysis of such similar phenotypes be carried out relative to

the cognate wild-type allele. It is also good practice to compare a mutant phenotype to a “null” allele (which today is typically created by carrying out a precise deletion of the entire gene).

Finally, much of human genetics focuses on mutations in human genes. As with mutations in the genes of experimental organisms, mutant alleles of some human genes can vary widely in phenotype. For example, more than half of all human tumors have mutations in the *TP53* gene. These are so-called “somatic mutations” that have arisen during the lifetime of the individual (as opposed to being inherited). Hundreds of different mutant *TP53* alleles have been found, ranging from simple changes in one base pair to deletions of the whole gene. These mutations produce diverse phenotypic effects, reflecting the diversity in the functions of the master regulatory protein that is encoded by the *TP53* gene. In addition, a few heritable DNA polymorphisms in the *TP53* gene are common in human populations for which no phenotype has been found. Other heritable *TP53* mutations cause a hereditary cancer predisposition disorder called Li–Fraumeni syndrome.

Accepting that one has to be sensitive about calling real people “mutants,” in reality the scientific issues that concern genetic mutations in humans are identical to those in experimental organisms; therefore, using the same scientific language when discussing them is entirely reasonable.

For further information on human genetic mutations, I refer readers to several important databases on human mutations that are accessible through the Internet, where you will find abundant references to the human genetic mutations mentioned in this book, as well as to the relevant scientific literature.

Databases on human mutations

- Human Gene Mutation Database (HGMD): <http://www.hgmd.org>
- Online Mendelian Inheritance in Man (OMIM): <http://www.ncbi.nlm.nih.gov/omim/>
- Genecards: <http://www.genecards.org/>
- Swiss Prot Diseases and Variants: <http://swissvar.expasy.org>

INTRODUCTORY BIOGRAPHIES

Theodosius Dobzhansky (1900–1975) was an eminent geneticist and evolutionary biologist. In addition to a prolific research career in *Drosophila* genetics, he wrote several influential articles and books that brought together genetics and evolutionary biology, which at the time were regarded as separate disciplines.

Gregor Mendel (1822–1884) was the founder of the science of genetics. In 1865–1866, he published an account of his experiments with pea plants that convincingly demonstrated the basic laws of inheritance. His achievement was not recognized until 1900, when three geneticists (Carl Erich Correns [1864–1933], Erich Tschermak von Seysenegg [1871–1962], and Hugo de Vries [1848–1935]) reintroduced his work, each having rediscovered some of Mendel’s findings.

Wilhelm Johannsen (1857–1927) was a Danish professor of botany. His interest in understanding the relationship between genetic and environmental causes of variation led him to coin the words “genotype” and “phenotype.” He wrote a very influential textbook in which he first used the word “gene.”

William Bateson (1861–1926) translated Mendel’s foundational paper into English and became the leading advocate for Mendelism in England. He coined the word “genetics” and wrote the first textbook on the subject. He also introduced the word “epistasis” to describe the masking of the phenotype of a mutation in one gene by a second mutation in another. With Reginald Punnett, he discovered genetic linkage, although he did not accept the then nascent chromosome theory.

Hermann J. Muller (1890–1967) discovered the mutagenic effects of X rays in *Drosophila* using an elegant genetic technique to detect lethal mutations. As a student at Columbia University, Muller had participated in the early development of *Drosophila* genetics in the remarkable group led by Thomas Hunt Morgan. He was active at the interface of genetics and society. In his later years, he became a leading antinuclear activist warning of the threat of nuclear war and weapons testing.

Hugo de Vries (1848–1935) was a professor of botany in Amsterdam who observed the 3:1 Mendelian ratio in his own experiments in Amsterdam in the late 19th century, which fueled the rediscovery of Mendel’s work in 1900. He gave the name “mutation” to suddenly appearing heritable variation.

Charles Darwin (1809–1882) was a British naturalist who introduced the most basic of biological principles: that species evolved from a common ancestor by “natural selection” of ever-fitter variants.