

MUTATIONS

Mutation

Alteration in the nucleotide sequence of a genome

This article is about the biological term. For other uses, see Mutation (disambiguation). In biology, a mutation is an alteration in the nucleic acid sequence of the genome of an organism, virus, or extrachromosomal DNA. Viral genomes contain either DNA or RNA. Mutations result from errors during DNA or viral replication, mitosis, or meiosis or other types of damage to DNA (such as pyrimidine dimers caused by exposure to ultraviolet radiation), which then may undergo error-prone repair (especially microhomology-mediated end joining), cause an error during other forms of repair, or cause an error during replication (translesion synthesis). Mutations may also result from substitution, insertion or deletion of segments of DNA due to mobile genetic elements. Mutations may or may not produce detectable changes in the observable characteristics (phenotype) of an organism. Mutations play a part in both normal and abnormal biological processes including: evolution, cancer, and the development of the immune system, including junctional diversity. Mutation is the ultimate source of all genetic variation, providing the raw material on which evolutionary forces such as natural selection can act.

*Mutation can result in many different types of change in sequences. Mutations in genes can have no effect, alter the product of a gene, or prevent the gene from functioning properly or completely. Mutations can also occur in non-genic regions. A 2007 study on genetic variations between different species of *Drosophila* suggested that, if a mutation changes a protein produced by a gene, the result is likely to be harmful, with an estimated 70% of amino acid polymorphisms that have damaging effects, and the remainder being either neutral or marginally beneficial.*

Mutation and DNA damage are the two major types of errors that occur in DNA, but they are fundamentally different. DNA damage is a physical alteration in the DNA structure, such as a single or double strand break, a modified guanosine residue in DNA such as 8-hydroxydeoxyguanosine, or a polycyclic aromatic hydrocarbon adduct. DNA damages can be recognized by enzymes, and therefore can be correctly repaired using the complementary undamaged strand in DNA as a template or an undamaged sequence in a homologous chromosome if it is available. If DNA damage remains in a cell, transcription of a gene may be prevented and thus translation into a protein may also be blocked. DNA replication may also be blocked and/or the cell may die. In contrast to a DNA damage, a mutation is an alteration of the base sequence of the DNA. Ordinarily, a mutation cannot be recognized by enzymes once the base change is present in both DNA strands, and thus a mutation is not ordinarily repaired. At the cellular level, mutations can alter protein function and regulation. Unlike DNA damages, mutations are replicated when the cell replicates. At the level of cell populations, cells with mutations will increase or decrease in frequency according to the effects of the mutations on the ability of the cell to survive and reproduce. Although distinctly different from each other, DNA damages and mutations are related because DNA damages often cause errors of DNA synthesis during replication or repair and these errors are a major source of mutation.

Overview

Mutations can involve the duplication of large sections of DNA, usually through genetic recombination. These duplications are a major source of raw material for evolving new genes, with tens to hundreds of genes duplicated in animal genomes every million years. Most genes belong to larger gene families of shared ancestry, detectable by their sequence homology. Novel genes are produced by several methods, commonly through the duplication and mutation of an ancestral gene, or by recombining parts of different genes to form new combinations with new functions.

Here, protein domains act as modules, each with a particular and independent function, that can be mixed together to produce genes encoding new proteins with novel properties. For example, the human eye uses four genes to make structures that sense light: three for cone cell or colour vision and one for rod cell or night vision; all four arose from a single ancestral gene. Another advantage

of duplicating a gene (or even an entire genome) is that this increases engineering redundancy; this allows one gene in the pair to acquire a new function while the other copy performs the original function. Other types of mutation occasionally create new genes from previously noncoding DNA.

Changes in chromosome number may involve even larger mutations, where segments of the DNA within chromosomes break and then rearrange. For example, in the Homininae, two chromosomes fused to produce human chromosome 2; this fusion did not occur in the lineage of the other apes, and they retain these separate chromosomes. In evolution, the most important role of such chromosomal rearrangements may be to accelerate the divergence of a population into new species by making populations less likely to interbreed, thereby preserving genetic differences between these populations.

Sequences of DNA that can move about the genome, such as transposons, make up a major fraction of the genetic material of plants and animals, and may have been important in the evolution of genomes. For example, more than a million copies of the Alu sequence are present in the human genome, and these sequences have now been recruited to perform functions such as regulating gene expression. Another effect of these mobile DNA sequences is that when they move within a genome, they can mutate or delete existing genes and thereby produce genetic diversity.

Nonlethal mutations accumulate within the gene pool and increase the amount of genetic variation. The abundance of some genetic changes within the gene pool can be reduced by natural selection, while other "more favorable" mutations may accumulate and result in adaptive changes.

For example, a butterfly may produce offspring with new mutations. The majority of these mutations will have no effect; but one might change the colour of one of the butterfly's offspring, making it harder (or easier) for predators to see. If this color change is advantageous, the chances of this butterfly's surviving and producing its own offspring are a little better, and over time the number of butterflies with this mutation may form a larger percentage of the population.[citation needed]

Neutral mutations are defined as mutations whose effects do not influence the fitness of an individual. These can increase in frequency over time due to genetic drift. It is believed that the overwhelming majority of mutations have no significant effect on an organism's fitness. Also, DNA repair mechanisms are able to mend most changes before they become permanent mutations, and many organisms have mechanisms, such as apoptotic pathways, for eliminating otherwise-permanently mutated somatic cells.

Beneficial mutations can improve reproductive success.

Causes

Mutagenesis

Four classes of mutations are (1) spontaneous mutations (molecular decay), (2) mutations due to error-prone replication bypass of naturally occurring DNA damage (also called error-prone translesion synthesis), (3) errors introduced during DNA repair, and (4) induced mutations caused by mutagens. Scientists may sometimes deliberately introduce mutations into cells or research organisms for the sake of scientific experimentation.

One 2017 study claimed that 66% of cancer-causing mutations are random, 29% are due to the environment (the studied population spanned 69 countries), and 5% are inherited.

Humans on average pass 60 new mutations to their children but fathers pass more mutations depending on their age with every year adding two new mutations to a child.

Spontaneous mutation

Spontaneous mutations occur with non-zero probability even given a healthy, uncontaminated cell. Naturally occurring oxidative DNA damage is estimated to occur 10,000 times per cell per day in humans and 100,000 times per cell per day in rats. Spontaneous mutations can be characterized by the specific change:

- *Tautomerism – A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base, resulting in incorrect base pairing during replication. Theoretical results suggest that proton tunnelling is an important factor in the spontaneous creation of GC tautomers.*
- *Depurination – Loss of a purine base (A or G) to form an apurinic site (AP site).*
- *Deamination – Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.*
- *Slipped strand mispairing – Denaturation of the new strand from the template during replication, followed by renaturation in a different spot ("slipping"). This can lead to insertions or deletions.*

Error-prone replication bypass

There is increasing evidence that the majority of spontaneously arising mutations are due to error-prone replication (translesion synthesis) past DNA damage in the template strand. In mice, the majority of mutations are caused by translesion synthesis. Likewise, in yeast, Kunz et al. found that more than 60% of the spontaneous single base pair substitutions and deletions were caused by translesion synthesis.

Errors introduced during DNA repair

See also: DNA damage (naturally occurring) and DNA repair

Although naturally occurring double-strand breaks occur at a relatively low frequency in DNA, their repair often causes mutation. Non-homologous end joining (NHEJ) is a major pathway for repairing double-strand breaks. NHEJ involves removal of a few nucleotides to allow somewhat inaccurate alignment of the two ends for rejoining followed by addition of nucleotides to fill in gaps. As a consequence, NHEJ often introduces mutations.

A covalent adduct between the metabolite of benzo[a]pyrene, the major mutagen in tobacco smoke, and DNA

Induced mutation

Induced mutations are alterations in the gene after it has come in contact with mutagens and environmental causes.

Induced mutations on the molecular level can be caused by:

❖ *Chemicals*

- *Hydroxylamine*
- *Base analogues (e.g., Bromodeoxyuridine (BrdU))*
- *Alkylating agents (e.g., N-ethyl-N-nitrosourea (ENU). These agents can mutate both replicating and non-replicating DNA. In contrast, a base analogue can mutate the DNA only when the analogue is incorporated in replicating the DNA. Each of these classes of chemical mutagens has certain effects that then lead to transitions, transversions, or deletions.*
- *Agents that form DNA adducts (e.g., ochratoxin A)*
- *DNA intercalating agents (e.g., ethidium bromide)*
- *DNA crosslinkers*
- *Oxidative damage*

- *Nitrous acid converts amine groups on A and C to diazo groups, altering their hydrogen bonding patterns, which leads to incorrect base pairing during replication.*

❖ *Radiation*

- *Ultraviolet light (UV) (including non-ionizing radiation). Two nucleotide bases in DNA—cytosine and thymine—are most vulnerable to radiation that can change their properties. UV light can induce adjacent pyrimidine bases in a DNA strand to become covalently joined as a pyrimidine dimer. UV radiation, in particular longer-wave UVA, can also cause oxidative damage to DNA.*
- *Ionizing radiation. Exposure to ionizing radiation, such as gamma radiation, can result in mutation, possibly resulting in cancer or death.*

Whereas in former times mutations were assumed to occur by chance, or induced by mutagens, molecular mechanisms of mutation have been discovered in bacteria and across the tree of life. As S. Rosenberg states, "These mechanisms reveal a picture of highly regulated mutagenesis, up-regulated temporally by stress responses and activated when cells/organisms are maladapted to their environments—when stressed—potentially accelerating adaptation." Since they are self-induced mutagenic mechanisms that increase the adaptation rate of organisms, they have some times been named as adaptive mutagenesis mechanisms, and include the SOS response in bacteria, ectopic intrachromosomal recombination and other chromosomal events such as duplications.

Classification of types

By effect on structure

The sequence of a gene can be altered in a number of ways. Gene mutations have varying effects on health depending on where they occur and whether they alter the function of essential proteins. Mutations in the structure of genes can be classified into several types.[citation needed]

Large-scale mutations

Large-scale mutations in chromosomal structure include:

- *Amplifications (or gene duplications) or repetition of a chromosomal segment or presence of extra piece of a chromosome broken piece of a chromosome may become attached to a homologous or non-homologous chromosome so that some of the genes are present in more than two doses leading to multiple copies of all chromosomal regions, increasing the dosage of the genes located within them.*
- *Polyploidy, duplication of entire sets of chromosomes, potentially resulting in a separate breeding population and speciation.*
- *Deletions of large chromosomal regions, leading to loss of the genes within those regions.*
- *Mutations whose effect is to juxtapose previously separate pieces of DNA, potentially bringing together separate genes to form functionally distinct fusion genes (e.g., bcr-abl).*
- *Large scale changes to the structure of chromosomes called chromosomal rearrangement that can lead to a decrease of fitness but also to speciation in isolated, inbred populations. These include:*
 - *Chromosomal translocations: interchange of genetic parts from nonhomologous chromosomes.*
 - *Chromosomal inversions: reversing the orientation of a chromosomal segment.*
 - *Non-homologous chromosomal crossover.*
 - *Interstitial deletions: an intra-chromosomal deletion that removes a segment of DNA from a single chromosome, thereby apposing previously distant genes. For example, cells isolated from a human astrocytoma, a type of brain tumour, were found to have a chromosomal deletion removing sequences between the Fused in Glioblastoma (FIG) gene and the receptor tyrosine kinase (ROS), producing a fusion protein (FIG-ROS). The abnormal FIG-ROS fusion protein has*

constitutively active kinase activity that causes oncogenic transformation (a transformation from normal cells to cancer cells).

- *Loss of heterozygosity: loss of one allele, either by a deletion or a genetic recombination event, in an organism that previously had two different alleles.*

Small-scale mutations

Small-scale mutations affect a gene in one or a few nucleotides. (If only a single nucleotide is affected, they are called point mutations.) Small-scale mutations include:

- I. Insertions add one or more extra nucleotides into the DNA. They are usually caused by transposable elements, or errors during replication of repeating elements. Insertions in the coding region of a gene may alter splicing of the mRNA (splice site mutation), or cause a shift in the reading frame (frameshift), both of which can significantly alter the gene product. Insertions can be reversed by excision of the transposable element.*
- II. Deletions remove one or more nucleotides from the DNA. Like insertions, these mutations can alter the reading frame of the gene. In general, they are irreversible: Though exactly the same sequence might, in theory, be restored by an insertion, transposable elements able to revert a very short deletion (say 1–2 bases) in any location either are highly unlikely to exist or do not exist at all.*
- III. Substitution mutations, often caused by chemicals or malfunction of DNA replication, exchange a single nucleotide for another. These changes are classified as transitions or transversions. Most common is the transition that exchanges a purine for a purine ($A \leftrightarrow G$) or a pyrimidine for a pyrimidine ($C \leftrightarrow T$). A transition can be caused by nitrous acid, base mispairing, or mutagenic base analogues such as BrdU. Less common is a transversion, which exchanges a purine for a pyrimidine or a pyrimidine for a purine ($C/T \leftrightarrow A/G$). An example of a transversion is the conversion of adenine (A) into a cytosine (C). Point mutations are modifications of single base pairs of DNA or other small base pairs within a gene. A point mutation can be reversed by another point mutation, in which the nucleotide is changed back to its original state (true reversion) or by second-site reversion (a complementary mutation elsewhere that results in regained gene functionality). As discussed below, point mutations that occur within the protein coding region of a gene may be classified as synonymous or nonsynonymous substitutions, the latter of which in turn can be divided into missense or nonsense mutations.*
- IV. By impact on protein sequence*

The effect of a mutation on protein sequence depends in part on where in the genome it occurs, especially whether it is in a coding or non-coding region. Mutations in the non-coding regulatory sequences of a gene, such as promoters, enhancers, and silencers, can alter levels of gene expression, but are less likely to alter the protein sequence. Mutations within introns and in regions with no known biological function (e.g. pseudogenes, retrotransposons) are generally neutral, having no effect on phenotype – though intron mutations could alter the protein product if they affect mRNA splicing.

Mutations that occur in coding regions of the genome are more likely to alter the protein product, and can be categorized by their effect on amino acid sequence:

- 1. A frameshift mutation is caused by insertion or deletion of a number of nucleotides that is not evenly divisible by three from a DNA sequence. Due to the triplet nature of gene expression by codons, the insertion or deletion can disrupt the reading frame, or the grouping of the codons, resulting in a completely different translation from the original. The earlier in the sequence the deletion or insertion occurs, the more altered the protein produced is. (For example, the code CCU GAC UAC CUA codes for the amino acids proline, aspartic acid, tyrosine, and leucine. If the U in CCU was deleted, the resulting sequence would be CCG ACU ACC UAx, which would instead code for proline, threonine, threonine, and part of another amino acid or perhaps a stop codon (where the x stands for*

the following nucleotide.) By contrast, any insertion or deletion that is evenly divisible by three is termed an in-frame mutation.

2. *A point substitution mutation results in a change in a single nucleotide and can be either synonymous or nonsynonymous.*
 - a. *A synonymous substitution replaces a codon with another codon that codes for the same amino acid, so that the produced amino acid sequence is not modified. Synonymous mutations occur due to the degenerate nature of the genetic code. If this mutation does not result in any phenotypic effects, then it is called silent, but not all synonymous substitutions are silent. (There can also be silent mutations in nucleotides outside of the coding regions, such as the introns, because the exact nucleotide sequence is not as crucial as it is in the coding regions, but these are not considered synonymous substitutions.)*
 - b. *A nonsynonymous substitution replaces a codon with another codon that codes for a different amino acid, so that the produced amino acid sequence is modified. Nonsynonymous substitutions can be classified as nonsense or missense mutations:*
 - i. *A missense mutation changes a nucleotide to cause substitution of a different amino acid. This in turn can render the resulting protein nonfunctional. Such mutations are responsible for diseases such as Epidermolysis bullosa, sickle-cell disease, and SOD1-mediated ALS. On the other hand, if a missense mutation occurs in an amino acid codon that results in the use of a different, but chemically similar, amino acid, then sometimes little or no change is rendered in the protein. For example, a change from AAA to AGA will encode arginine, a chemically similar molecule to the intended lysine. In this latter case the mutation will have little or no effect on phenotype and therefore be neutral.*
 - ii. *A nonsense mutation is a point mutation in a sequence of DNA that results in a premature stop codon, or a nonsense codon in the transcribed mRNA, and possibly a truncated, and often nonfunctional protein product. This sort of mutation has been linked to different diseases, such as congenital adrenal hyperplasia. (See Stop codon.)*

By effect on function

A mutation becomes an effect on function mutation when the exactitude of functions between a mutated protein and its direct interactor undergoes change. The interactors can be other proteins, molecules, nucleic acids, etc. There are many mutations that fall under the category of by effect on function, but depending on the specificity of the change the mutations listed below will occur.

- I. *Loss-of-function mutations, also called inactivating mutations, result in the gene product having less or no function (being partially or wholly inactivated). When the allele has a complete loss of function (null allele), it is often called an amorph or amorphic mutation in Muller's morphs schema. Phenotypes associated with such mutations are most often recessive. Exceptions are when the organism is haploid, or when the reduced dosage of a normal gene product is not enough for a normal phenotype (this is called haploinsufficiency). A disease that is caused by a loss-of-function mutation is Gitelman syndrome and cystic fibrosis.*
- II. *Gain-of-function mutations also called activating mutations, change the gene product such that its effect gets stronger (enhanced activation) or even is superseded by a different and abnormal function. When the new allele is created, a heterozygote containing the newly created allele as well as the original will express the new allele; genetically this defines the mutations as dominant phenotypes. Several of Muller's morphs correspond to the gain of function, including hypermorph (increased gene expression) and neomorph (novel function).*
- III. *Dominant negative mutations (also called anti-morphic mutations) have an altered gene product that acts antagonistically to the wild-type allele. These mutations usually result in an altered molecular function (often inactive) and are characterized by a dominant or*

semi-dominant phenotype. In humans, dominant negative mutations have been implicated in cancer (e.g., mutations in genes p53, ATM, CEBPA, and PPARGgamma). Marfan syndrome is caused by mutations in the FBN1 gene, located on chromosome 15, which encodes fibrillin-1, a glycoprotein component of the extracellular matrix. Marfan syndrome is also an example of dominant negative mutation and haploinsufficiency.

- IV. Lethal mutations result in rapid organismal death when occurring during development and cause significant reductions of life expectancy for developed organisms. An example of a disease that is caused by a dominant lethal mutation is Huntington's disease.*
- V. Null mutations, also known as Amorphic mutations, are a form of loss-of-function mutations that completely prohibit the gene's function. The mutation leads to a complete loss of operation at the phenotypic level, also causing no gene product to be formed. Atopic eczema and dermatitis syndrome are common diseases caused by a null mutation of the gene that activates filaggrin.*
- VI. Suppressor mutations are a type of mutation that causes the double mutation to appear normally. In suppressor mutations the phenotypic activity of a different mutation is completely suppressed, thus causing the double mutation to look normal. There are two types of suppressor mutations, there are intragenic and extragenic suppressor mutations. Intragenic mutations occur in the gene where the first mutation occurs, while extragenic mutations occur in the gene that interacts with the product of the first mutation. A common disease that results from this type of mutation is Alzheimer's disease.*
- VII. Neomorphic mutations are a part of the gain-of-function mutations and are characterized by the control of new protein product synthesis. The newly synthesized gene normally contains a novel gene expression or molecular function. The result of the neomorphic mutation is the gene where the mutation occurs has a complete change in function.*
- VIII. A back mutation or reversion is a point mutation that restores the original sequence and hence the original phenotype.*

By effect on fitness (harmful, beneficial, neutral mutations)

In genetics, it is sometimes useful to classify mutations as either harmful or beneficial (or neutral):

- I. A harmful, or deleterious, mutation decreases the fitness of the organism. Many, but not all mutations in essential genes are harmful (if a mutation does not change the amino acid sequence in an essential protein, it is harmless in most cases).*
- II. A beneficial, or advantageous mutation increases the fitness of the organism. Examples are mutations that lead to antibiotic resistance in bacteria (which are beneficial for bacteria but usually not for humans).*
- III. A neutral mutation has no harmful or beneficial effect on the organism. Such mutations occur at a steady rate, forming the basis for the molecular clock. In the neutral theory of molecular evolution, neutral mutations provide genetic drift as the basis for most variation at the molecular level. In animals or plants, most mutations are neutral, given that the vast majority of their genomes is either non-coding or consists of repetitive sequences that have no obvious function ("junk DNA").*

*Large-scale quantitative mutagenesis screens, in which thousands of millions of mutations are tested, invariably find that a larger fraction of mutations has harmful effects but always returns a number of beneficial mutations as well. For instance, in a screen of all gene deletions in *E. coli*, 80% of mutations were negative, but 20% were positive, even though many had a very small effect on growth (depending on condition). Gene deletions involve removal of whole genes, so that point mutations almost always have a much smaller effect. In a similar screen in *Streptococcus pneumoniae*, but this time with transposon insertions, 76% of insertion mutants were classified as neutral, 16% had a significantly reduced fitness, but 6% were advantageous.*

This classification is obviously relative and somewhat artificial: a harmful mutation can quickly turn into a beneficial mutations when conditions change. Also, there is a gradient from harmful/beneficial to neutral, as many mutations may have small and mostly neglectable effects

but under certain conditions will become relevant. Also, many traits are determined by hundreds of genes (or loci), so that each locus has only a minor effect. For instance, human height is determined by hundreds of genetic variants ("mutations") but each of them has a very minor effect on height, apart from the impact of nutrition. Height (or size) itself may be more or less beneficial as the huge range of sizes in animal or plant groups shows.

Distribution of fitness effects (DFE)

Attempts have been made to infer the distribution of fitness effects (DFE) using mutagenesis experiments and theoretical models applied to molecular sequence data. DFE, as used to determine the relative abundance of different types of mutations (i.e., strongly deleterious, nearly neutral or advantageous), is relevant to many evolutionary questions, such as the maintenance of genetic variation, the rate of genomic decay, the maintenance of outcrossing sexual reproduction as opposed to inbreeding and the evolution of sex and genetic recombination. DFE can also be tracked by tracking the skewness of the distribution of mutations with putatively severe effects as compared to the distribution of mutations with putatively mild or absent effect. In summary, the DFE plays an important role in predicting evolutionary dynamics. A variety of approaches have been used to study the DFE, including theoretical, experimental and analytical methods.

- Mutagenesis experiment:** The direct method to investigate the DFE is to induce mutations and then measure the mutational fitness effects, which has already been done in viruses, bacteria, yeast, and *Drosophila*. For example, most studies of the DFE in viruses used site-directed mutagenesis to create point mutations and measure relative fitness of each mutant. In *Escherichia coli*, one study used transposon mutagenesis to directly measure the fitness of a random insertion of a derivative of Tn10. In yeast, a combined mutagenesis and deep sequencing approach has been developed to generate high-quality systematic mutant libraries and measure fitness in high throughput. However, given that many mutations have effects too small to be detected and that mutagenesis experiments can detect only mutations of moderately large effect; DNA sequence analysis can provide valuable information about these mutations. **Molecular sequence analysis:** With rapid development of DNA sequencing technology, an enormous amount of DNA sequence data is available and even more is forthcoming in the future. Various methods have been developed to infer the DFE from DNA sequence data. By examining DNA sequence differences within and between species, we are able to infer various characteristics of the DFE for neutral, deleterious and advantageous mutations. To be specific, the DNA sequence analysis approach allows us to estimate the effects of mutations with very small effects, which are hardly detectable through mutagenesis experiments.

One of the earliest theoretical studies of the distribution of fitness effects was done by Motoo Kimura, an influential theoretical population geneticist. His neutral theory of molecular evolution proposes that most novel mutations will be highly deleterious, with a small fraction being neutral. A later proposal by Hiroshi Akashi proposed a bimodal model for the DFE, with modes centered around highly deleterious and neutral mutations. Both theories agree that the vast majority of novel mutations are neutral or deleterious and that advantageous mutations are rare, which has been supported by experimental results. One example is a study done on the DFE of random mutations in vesicular stomatitis virus. Out of all mutations, 39.6% were lethal, 31.2% were non-lethal deleterious, and 27.1% were neutral. Another example comes from a high throughput mutagenesis experiment with yeast. In this experiment it was shown that the overall DFE is bimodal, with a cluster of neutral mutations, and a broad distribution of deleterious mutations.

Though relatively few mutations are advantageous, those that are play an important role in evolutionary changes. Like neutral mutations, weakly selected advantageous mutations can be lost due to random genetic drift, but strongly selected advantageous mutations are more likely to be fixed. Knowing the DFE of advantageous mutations may lead to increased ability to predict the evolutionary dynamics. Theoretical work on the DFE for advantageous mutations has been done by John H. Gillespie and H. Allen Orr. They proposed that the distribution for advantageous

mutations should be exponential under a wide range of conditions, which, in general, has been supported by experimental studies, at least for strongly selected advantageous mutations.

In general, it is accepted that the majority of mutations are neutral or deleterious, with advantageous mutations being rare; however, the proportion of types of mutations varies between species. This indicates two important points: first, the proportion of effectively neutral mutations is likely to vary between species, resulting from dependence on effective population size; second, the average effect of deleterious mutations varies dramatically between species. In addition, the DFE also differs between coding regions and noncoding regions, with the DFE of noncoding DNA containing more weakly selected mutations.

By inheritance

In multicellular organisms with dedicated reproductive cells, mutations can be subdivided into germline mutations, which can be passed on to descendants through their reproductive cells, and somatic mutations (also called acquired mutations), which involve cells outside the dedicated reproductive group and which are not usually transmitted to descendants.

Diploid organisms (e.g., humans) contain two copies of each gene—a paternal and a maternal allele. Based on the occurrence of mutation on each chromosome, we may classify mutations into three types. A wild type or homozygous non-mutated organism is one in which neither allele is mutated.

- 1. A heterozygous mutation is a mutation of only one allele.*
- 2. A homozygous mutation is an identical mutation of both the paternal and maternal alleles.*
- 3. Compound heterozygous mutations or a genetic compound consists of two different mutations in the paternal and maternal alleles.*

Germline mutation

A germline mutation in the reproductive cells of an individual gives rise to a constitutional mutation in the offspring, that is, a mutation that is present in every cell. A constitutional mutation can also occur very soon after fertilization, or continue from a previous constitutional mutation in a parent. A germline mutation can be passed down through subsequent generations of organisms.

The distinction between germline and somatic mutations is important in animals that have a dedicated germline to produce reproductive cells. However, it is of little value in understanding the effects of mutations in plants, which lack a dedicated germline. The distinction is also blurred in those animals that reproduce asexually through mechanisms such as budding, because the cells that give rise to the daughter organisms also give rise to that organism's germline.

A new germline mutation not inherited from either parent is called a de novo mutation.

Somatic mutation

A change in the genetic structure that is not inherited from a parent, and also not passed to offspring, is called a somatic mutation. Somatic mutations are not inherited by an organism's offspring because they do not affect the germline. However, they are passed down to all the progeny of a mutated cell within the same organism during mitosis. A major section of an organism therefore might carry the same mutation. These types of mutations are usually prompted by environmental causes, such as ultraviolet radiation or any exposure to certain harmful chemicals, and can cause diseases including cancer.

With plants, some somatic mutations can be propagated without the need for seed production, for example, by grafting and stem cuttings. These type of mutation have led to new types of fruits, such as the "Delicious" apple and the "Washington" navel orange.

Human and mouse somatic cells have a mutation rate more than ten times higher than the germline mutation rate for both species; mice have a higher rate of both somatic and germline mutations per cell division than humans. The disparity in mutation rate between the germline and somatic tissues likely reflects the greater importance of genome maintenance in the germline than in the soma.

Special classes

1. *Conditional mutation is a mutation that has wild-type (or less severe) phenotype under certain "permissive" environmental conditions and a mutant phenotype under certain "restrictive" conditions. For example, a temperature-sensitive mutation can cause cell death at high temperature (restrictive condition), but might have no deleterious consequences at a lower temperature (permissive condition). These mutations are non-autonomous, as their manifestation depends upon presence of certain conditions, as opposed to other mutations which appear autonomously. The permissive conditions may be temperature, certain chemicals, light or mutations in other parts of the genome. In vivo mechanisms like transcriptional switches can create conditional mutations. For instance, association of Steroid Binding Domain can create a transcriptional switch that can change the expression of a gene based on the presence of a steroid ligand. Conditional mutations have applications in research as they allow control over gene expression. This is especially useful studying diseases in adults by allowing expression after a certain period of growth, thus eliminating the deleterious effect of gene expression seen during stages of development in model organisms. DNA Recombinase systems like Cre-Lox recombination used in association with promoters that are activated under certain conditions can generate conditional mutations. Dual Recombinase technology can be used to induce multiple conditional mutations to study the diseases which manifest as a result of simultaneous mutations in multiple genes. Certain inteins have been identified which splice only at certain permissive temperatures, leading to improper protein synthesis and thus, loss-of-function mutations at other temperatures. Conditional mutations may also be used in genetic studies associated with ageing, as the expression can be changed after a certain time period in the organism's lifespan.*
2. *Replication timing quantitative trait loci affects DNA replication.*

Nomenclature

In order to categorize a mutation as such, the "normal" sequence must be obtained from the DNA of a "normal" or "healthy" organism (as opposed to a "mutant" or "sick" one), it should be identified and reported; ideally, it should be made publicly available for a straightforward nucleotide-by-nucleotide comparison, and agreed upon by the scientific community or by a group of expert geneticists and biologists, who have the responsibility of establishing the standard or so-called "consensus" sequence. This step requires a tremendous scientific effort. Once the consensus sequence is known, the mutations in a genome can be pinpointed, described, and classified. The committee of the Human Genome Variation Society (HGVS) has developed the standard human sequence variant nomenclature, which should be used by researchers and DNA diagnostic centers to generate unambiguous mutation descriptions. In principle, this nomenclature can also be used to describe mutations in other organisms. The nomenclature specifies the type of mutation and base or amino acid changes.

- I. *Nucleotide substitution (e.g., 76A>T) – The number is the position of the nucleotide from the 5' end; the first letter represents the wild-type nucleotide, and the second letter represents the nucleotide that replaced the wild type. In the given example, the adenine at the 76th position was replaced by a thymine.*
 - A. *If it becomes necessary to differentiate between mutations in genomic DNA, mitochondrial DNA, and RNA, a simple convention is used. For example, if the 100th base of a nucleotide sequence mutated from G to C, then it would be written as g.100G>C if the mutation occurred in genomic DNA, m.100G>C if the mutation*

occurred in mitochondrial DNA, or r.100g>c if the mutation occurred in RNA.

Note that, for mutations in RNA, the nucleotide code is written in lower case.

- II. Amino acid substitution (e.g., D111E) – The first letter is the one letter code of the wild-type amino acid, the number is the position of the amino acid from the N-terminus, and the second letter is the one letter code of the amino acid present in the mutation. Nonsense mutations are represented with an X for the second amino acid (e.g. D111X).*
- III. Amino acid deletion (e.g., ΔF508) – The Greek letter Δ (delta) indicates a deletion. The letter refers to the amino acid present in the wild type and the number is the position from the N terminus of the amino acid were it to be present as in the wild type.*

Mutation rates

Mutation rates vary substantially across species, and the evolutionary forces that generally determine mutation are the subject of ongoing investigation.

In humans, the mutation rate is about 50–90 de novo mutations per genome per generation, that is, each human accumulates about 50–90 novel mutations that were not present in his or her parents. This number has been established by sequencing thousands of human trios, that is, two parents and at least one child.

The genomes of RNA viruses are based on RNA rather than DNA. The RNA viral genome can be double-stranded (as in DNA) or single-stranded. In some of these viruses (such as the single-stranded human immunodeficiency virus), replication occurs quickly, and there are no mechanisms to check the genome for accuracy. This error-prone process often results in mutations.

The rate of de novo mutations, whether germline or somatic, vary among organisms. Individuals within the same species can even express varying rates of mutation. Overall, rates of de novo mutations are low compared to those of inherited mutations, which categorizes them as rare forms of genetic variation. Many observations of de novo mutation rates have associated higher rates of mutation correlated to paternal age. In sexually reproducing organisms, the comparatively higher frequency of cell divisions in the parental sperm donor germline drive conclusions that rates of de novo mutation can be tracked along a common basis. The frequency of error during the DNA replication process of gametogenesis, especially amplified in the rapid production of sperm cells, can promote more opportunities for de novo mutations to replicate unregulated by DNA repair machinery. This claim combines the observed effects of increased probability for mutation in rapid spermatogenesis with short periods of time between cellular divisions that limit the efficiency of repair machinery. Rates of de novo mutations that affect an organism during its development can also increase with certain environmental factors. For example, certain intensities of exposure to radioactive elements can inflict damage to an organism's genome, heightening rates of mutation. In humans, the appearance of skin cancer during one's lifetime is induced by overexposure to UV radiation that causes mutations in the cellular and skin genome.

Randomness of mutations

*There is a widespread assumption that mutations are (entirely) "random" with respect to their consequences (in terms of probability). This was shown to be wrong as mutation frequency can vary across regions of the genome, with such DNA repair- and mutation-biases being associated with various factors. For instance, Monroe and colleagues demonstrated that—in the studied plant (*Arabidopsis thaliana*)—more important genes mutate less frequently than less important ones. They demonstrated that mutation is "non-random in a way that benefits the plant". Additionally, previous experiments typically used to demonstrate mutations being random with respect to fitness (such as the Fluctuation Test and Replica plating) have been shown to only support the weaker claim that those mutations are random with respect to external selective constraints, not fitness as a whole.*

Disease causation

*Changes in DNA caused by mutation in a coding region of DNA can cause errors in protein sequence that may result in partially or completely non-functional proteins. Each cell, in order to function correctly, depends on thousands of proteins to function in the right places at the right times. When a mutation alters a protein that plays a critical role in the body, a medical condition can result. One study on the comparison of genes between different species of *Drosophila* suggests that if a mutation does change a protein, the mutation will most likely be harmful, with an estimated 70 per cent of amino acid polymorphisms having damaging effects, and the remainder being either neutral or weakly beneficial. Some mutations alter a gene's DNA base sequence but do not change the protein made by the gene. Studies have shown that only 7% of point mutations in noncoding DNA of yeast are deleterious and 12% in coding DNA are deleterious. The rest of the mutations are either neutral or slightly beneficial.*

Inherited disorders

If a mutation is present in a germ cell, it can give rise to offspring that carries the mutation in all of its cells. This is the case in hereditary diseases. In particular, if there is a mutation in a DNA repair gene within a germ cell, humans carrying such germline mutations may have an increased risk of cancer. A list of 34 such germline mutations is given in the article DNA repair-deficiency disorder. An example of one is albinism, a mutation that occurs in the OCA1 or OCA2 gene. Individuals with this disorder are more prone to many types of cancers, other disorders and have impaired vision.

DNA damage can cause an error when the DNA is replicated, and this error of replication can cause a gene mutation that, in turn, could cause a genetic disorder. DNA damages are repaired by the DNA repair system of the cell. Each cell has a number of pathways through which enzymes recognize and repair damages in DNA. Because DNA can be damaged in many ways, the process of DNA repair is an important way in which the body protects itself from disease. Once DNA damage has given rise to a mutation, the mutation cannot be repaired.

Role in carcinogenesis

On the other hand, a mutation may occur in a somatic cell of an organism. Such mutations will be present in all descendants of this cell within the same organism. The accumulation of certain mutations over generations of somatic cells is part of cause of malignant transformation, from normal cell to cancer cell.

Cells with heterozygous loss-of-function mutations (one good copy of gene and one mutated copy) may function normally with the unmutated copy until the good copy has been spontaneously somatically mutated. This kind of mutation happens often in living organisms, but it is difficult to measure the rate. Measuring this rate is important in predicting the rate at which people may develop cancer.

Point mutations may arise from spontaneous mutations that occur during DNA replication. The rate of mutation may be increased by mutagens. Mutagens can be physical, such as radiation from UV rays, X-rays or extreme heat, or chemical (molecules that misplace base pairs or disrupt the helical shape of DNA). Mutagens associated with cancers are often studied to learn about cancer and its prevention.

Beneficial and conditional mutations

Although mutations that cause changes in protein sequences can be harmful to an organism, on occasions the effect may be positive in a given environment. In this case, the mutation may enable the mutant organism to withstand particular environmental stresses better than wild-type organisms, or reproduce more quickly. In these cases a mutation will tend to become more common in a population through natural selection. That said, the same mutation can be beneficial in one condition and disadvantageous in another condition. Examples include the following:

- I. HIV resistance: a specific 32 base pair deletion in human CCR5 (CCR5-Δ32) confers HIV resistance to homozygotes and delays AIDS onset in heterozygotes. One possible***

explanation of the etiology of the relatively high frequency of CCR5-Δ32 in the European population is that it conferred resistance to the bubonic plague in mid-14th century Europe. People with this mutation were more likely to survive infection; thus its frequency in the population increased. This theory could explain why this mutation is not found in Southern Africa, which remained untouched by bubonic plague. A newer theory suggests that the selective pressure on the CCR5 Delta 32 mutation was caused by smallpox instead of the bubonic plague.

- II. Malaria resistance: An example of a harmful mutation is sickle-cell disease, a blood disorder in which the body produces an abnormal type of the oxygen-carrying substance haemoglobin in the red blood cells. One-third of all indigenous inhabitants of Sub-Saharan Africa carry the allele, because, in areas where malaria is common, there is a survival value in carrying only a single sickle-cell allele (sickle cell trait). Those with only one of the two alleles of the sickle-cell disease are more resistant to malaria, since the infestation of the malaria Plasmodium is halted by the sickling of the cells that it infests.*
- III. Antibiotic resistance: Practically all bacteria develop antibiotic resistance when exposed to antibiotics. In fact, bacterial populations already have such mutations that get selected under antibiotic selection. Obviously, such mutations are only beneficial for the bacteria but not for those infected.*
- IV. Lactase persistence. A mutation allowed humans to express the enzyme lactase after they are naturally weaned from breast milk, allowing adults to digest lactose, which is likely one of the most beneficial mutations in recent human evolution.*

Role in evolution

By introducing novel genetic qualities to a population of organisms, de novo mutations play a critical role in the combined forces of evolutionary change. However, the weight of genetic diversity generated by mutational change is often considered a generally "weak" evolutionary force. Although the random emergence of mutations alone provides the basis for genetic variation across all organic life, this force must be taken in consideration alongside all evolutionary forces at play. Spontaneous de novo mutations as cataclysmic events of speciation depend on factors introduced by natural selection, genetic flow, and genetic drift. For example, smaller populations with heavy mutational input (high rates of mutation) are prone to increases of genetic variation which lead to speciation in future generations. In contrast, larger populations tend to see lesser effects of newly introduced mutated traits. In these conditions, selective forces diminish the frequency of mutated alleles, which are most often deleterious, over time.

Compensated pathogenic deviations

Compensated pathogenic deviations refer to amino acid residues in a protein sequence that are pathogenic in one species but are wild type residues in the functionally equivalent protein in another species. Although the amino acid residue is pathogenic in the first species, it is not so in the second species because its pathogenicity is compensated by one or more amino acid substitutions in the second species. The compensatory mutation can occur in the same protein or in another protein with which it interacts.

It is critical to understand the effects of compensatory mutations in the context of fixed deleterious mutations due to the population fitness decreasing because of fixation. Effective population size refers to a population that is reproducing. An increase in this population size has been correlated with a decreased rate of genetic diversity. The position of a population relative to the critical effect population size is essential to determine the effect deleterious alleles will have on fitness. If the population is below the critical effective size fitness will decrease drastically, however if the population is above the critical effect size, fitness can increase regardless of deleterious mutations due to compensatory alleles.

Compensatory mutations in RNA

As the function of a RNA molecule is dependent on its structure, the structure of RNA molecules is evolutionarily conserved. Therefore, any mutation that alters the stable structure of RNA molecules must be compensated by other compensatory mutations. In the context of RNA, the sequence of the RNA can be considered as 'genotype' and the structure of the RNA can be considered as its 'phenotype'. Since RNAs have relatively simpler composition than proteins, the structure of RNA molecules can be computationally predicted with high degree of accuracy. Because of this convenience, compensatory mutations have been studied in computational simulations using RNA folding algorithms.

Evolutionary mechanism of compensation

Compensatory mutations can be explained by the genetic phenomenon epistasis whereby the phenotypic effect of one mutation is dependent upon mutation(s) at other loci. While epistasis was originally conceived in the context of interaction between different genes, intragenic epistasis has also been studied recently. Existence of compensated pathogenic deviations can be explained by 'sign epistasis', in which the effects of a deleterious mutation can be compensated by the presence of an epistatic mutation in another loci. For a given protein, a deleterious mutation (D) and a compensatory mutation (C) can be considered, where C can be in the same protein as D or in a different interacting protein depending on the context. The fitness effect of C itself could be neutral or somewhat deleterious such that it can still exist in the population, and the effect of D is deleterious to the extent that it cannot exist in the population. However, when C and D co-occur together, the combined fitness effect becomes neutral or positive. Thus, compensatory mutations can bring novelty to proteins by forging new pathways of protein evolution : it allows individuals to travel from one fitness peak to another through the valleys of lower fitness.

DePristo et al. 2005 outlined two models to explain the dynamics of compensatory pathogenic deviations (CPD). In the first hypothesis P is a pathogenic amino acid mutation that and C is a neutral compensatory mutation. Under these conditions, if the pathogenic mutation arises after a compensatory mutation, then P can become fixed in the population. The second model of CPDs states that P and C are both deleterious mutations resulting in fitness valleys when mutations occur simultaneously. Using publicly available, Ferrer-Costa et al. 2007 obtained compensatory mutations and human pathogenic mutation datasets that were characterized to determine what causes CPDs. Results indicate that the structural constraints and the location in protein structure determine whether compensated mutations will occur.

Experimental evidence of compensatory mutations

Experiment in bacteria

Lunzer et al. tested the outcome of swapping divergent amino acids between two orthologous proteins of isopropylmalate dehydrogenase (IMDH). They substituted 168 amino acids in Escherichia coli IMDH that are wild type residues in IMDH Pseudomonas aeruginosa. They found that over one third of these substitutions compromised IMDH enzymatic activity in the Escherichia coli genetic background. This demonstrated that identical amino acid states can result in different phenotypic states depending on the genetic background. Corrigan et al. 2011 demonstrated how Staphylococcus aureus was able to grow normally without the presence of lipoteichoic acid due to compensatory mutations. Whole genome sequencing results revealed that when Cyclic-di-AMP phosphodiesterase (GdpP) was disrupted in this bacterium, it compensated for the disappearance of the cell wall polymer, resulting in normal cell growth.

Research has shown that bacteria can gain drug resistance through compensatory mutations that do not impede or having little effect on fitness. Previous research from Gagneux et al. 2006 has found that laboratory grown Mycobacterium tuberculosis strains with rifampicin resistance have reduced fitness, however drug resistant clinical strains of this pathogenic bacteria do not have reduced fitness. Comas et al. 2012 used whole genome comparisons between clinical strains and lab derived mutants to determine the role and contribution of compensatory mutations in drug resistance to rifampicin. Genome analysis reveal rifampicin resistant strains have a mutation in

rpoA and *rpoC*. A similar study investigated the bacterial fitness associated with compensatory mutations in rifampin resistant *Escherichia coli*. Results obtained from this study demonstrate that drug resistance is linked to bacterial fitness as higher fitness costs are linked to greater transcription errors.

Experiment in virus

Gong et al. collected obtained genotype data of influenza nucleoprotein from different timelines and temporally ordered them according to their time of origin. Then they isolated 39 amino acid substitutions that occurred in different timelines and substituted them in a genetic background that approximated the ancestral genotype. They found that 3 of the 39 substitutions significantly reduced the fitness of the ancestral background. Compensatory mutations are new mutations that arise and have a positive or neutral impact on a population's fitness. Previous research has shown that populations can compensate detrimental mutations. Burch and Chao tested Fisher's geometric model of adaptive evolution by testing whether bacteriophage $\phi 6$ evolves by small steps. Their results showed that bacteriophage $\phi 6$ fitness declined rapidly and recovered in small steps. Viral nucleoproteins have been shown to avoid cytotoxic T lymphocytes (CTLs) through arginine-to-glycine substitutions. This substitution impacts the fitness of viral nucleoproteins, however compensatory co-mutations impede fitness declines and aid the virus to avoid recognition from CTLs. Mutations can have three different effects; mutations can have deleterious effects, some increase fitness through compensatory mutations, and lastly mutations can be counterbalancing resulting in compensatory neutral mutations.

Application in human evolution and disease

In the human genome, the frequency and characteristics of *de novo* mutations have been studied as important contextual factors to our evolution. Compared to the human reference genome, a typical human genome varies at approximately 4.1 to 5.0 million loci, and the majority of this genetic diversity is shared by nearly 0.5% of the population. The typical human genome also contains 40,000 to 200,000 rare variants observed in less than 0.5% of the population that can only have occurred from at least one *de novo* germline mutation in the history of human evolution. *De novo* mutations have also been researched as playing a crucial role in the persistence of genetic disease in humans. With recent advancements in next-generation sequencing (NGS), all types of *de novo* mutations within the genome can be directly studied, the detection of which provides a magnitude of insight toward the causes of both rare and common genetic disorders. Currently, the best estimate of the average human germline SNV mutation rate is 1.18×10^{-8} , with an approximate ~78 novel mutations per generation. The ability to conduct whole genome sequencing of parents and offspring allows for the comparison of mutation rates between generations, narrowing down the origin possibilities of certain genetic disorders.

Related Extensions

- I. Aneuploidy
- II. Antioxidant
- III. Behavior mutation
- IV. Budgerigar colour genetics
- V. DbDNV (2010)
- VI. Deletion (genetics)
- VII. Ecogenetics
- VIII. Embryology
- IX. Homeobox
- X. Human somatic variation
- XI. Polyploidy
- XII. Robertsonian translocation
- XIII. Signature-tagged mutagenesis
- XIV. Somatic hypermutation
- XV. TILLING (molecular biology)
- XVI. Trinucleotide repeat expansion

