Mutations

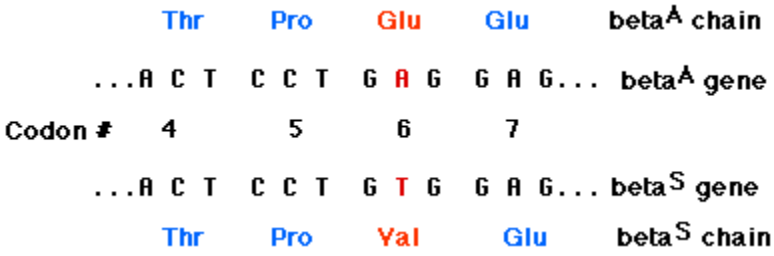
In the living cell, DNA undergoes frequent chemical change, especially when it is being replicated (in S phase of the eukaryotic cell cycle). Most of these changes are quickly repaired. Those that are not result in a mutation. Thus, **mutation is a failure of DNA repair**.

Single-base substitutions

A single base, say an A, becomes replaced by another. Single base substitutions are also called **point mutations**. (If one purine [A or G] or pyrimidine [C or T] is replaced by the other, the substitution is called a **transition**. If a purine is replaced by a pyrimidine or vice-versa, the substitution is called a **transversion**.)

Missense mutations

With a missense mutation, the new nucleotide alters the codon so as to produce an altered amino acid in the protein product.

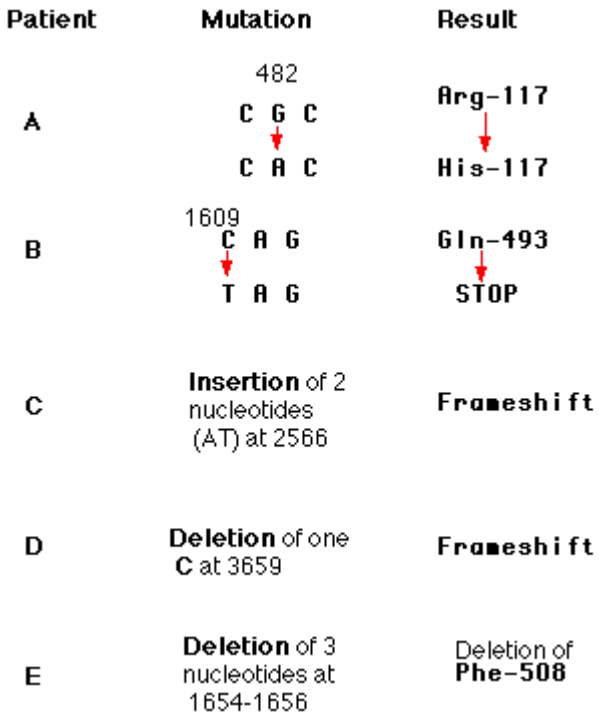


EXAMPLE: **sickle-cell disease** The replacement of A by T at the 17th nucleotide of the gene for the beta chain of hemoglobin changes the codon GAG (for glutamic acid) to GTG (which encodes valine). Thus the 6th amino acid in the chain becomes valine instead of glutamic acid.

ANOTHER EXAMPLE: **Patient A** with **cystic fibrosis** (scroll down).

Nonsense mutations

With a nonsense mutation, the new nucleotide changes a codon that specified an amino acid to one of the **STOP** codons (**TAA**, **TAG**, or **TGA**). Therefore, **translation** of the messenger RNA **transcribed** from this mutant gene will stop prematurely. The earlier in the gene that this occurs, the more truncated the protein product and the more likely that it will be unable to function.



EXAMPLE: **Patient B**

Here is a sampling of the more than 1000 different mutations that have been found in patients with **cystic fibrosis**. Each of these mutations occurs in a huge gene that encodes a protein (of 1480 amino acids) called the **c**ystic **f**ibrosis **t**ransmembrane conductance **r**egulator (**CFTR**). The protein is responsible for transporting chloride ions through the plasma membrane. The gene encompasses over 6000 nucleotides spread over 27 exons on chromosome 7. The numbers in the mutation column represent the number of the nucleotides affected. Defects in the protein cause the various symptoms of the disease. Unlike sickle-cell disease, then, no single mutation is responsible for all cases of cystic fibrosis. People with cystic fibrosis inherit two mutant genes, but the mutations need not be the same.

In one patient with cystic fibrosis (**Patient B**), the substitution of a T for a C at nucleotide 1609 converted a glutamine codon (**CAG**) to a **STOP** codon (**TAG**). The protein produced by this patient had only the first 493 amino acids of the normal chain of 1480 and could not function.

Silent mutations

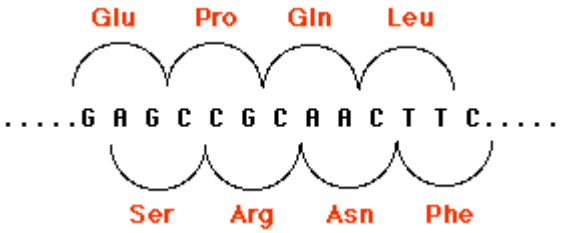
Most amino acids are encoded by several different codons. For example, if the third base in the **TCT** codon for **serine** is changed to any one of the other three bases, serine will still be encoded. Such mutations are said to be silent because they cause no change in their product and cannot be detected without sequencing the gene (or its mRNA).

Splice-site mutations

The removal of intron sequences, as pre-mRNA is being processed to form mRNA, must be done with great precision. Nucleotide signals at the splice sites guide the enzymatic machinery. If a mutation alters one of these signals, then the intron is not removed and remains as part of the final RNA molecule. The translation of its sequence alters the sequence of the protein product.

Insertions and Deletions (Indels)

Extra base pairs may be added (**insertions**) or removed (**deletions**) from the DNA of a gene. The number can range from one to thousands. Collectively, these mutations are called **indels**.



Indels involving one or two base pairs (or multiples thereof) can have devastating consequences to the gene because translation of the gene is "frameshifted". This figure shows how by **shifting the reading frame** one nucleotide to the right, the same sequence of nucleotides encodes a different sequence of amino acids. The mRNA is translated in new groups of three nucleotides and the protein specified by these new codons will be worthless. Scroll up to see two other examples (**Patients C and D**).

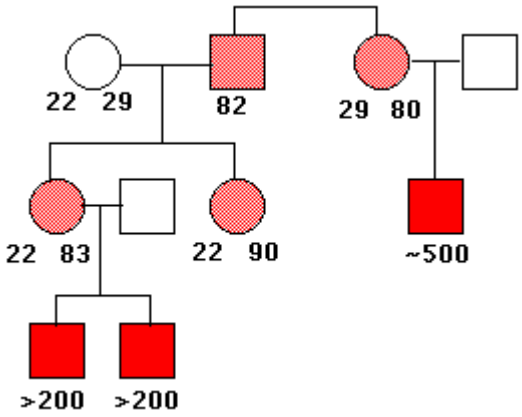
Frameshifts often create new **STOP** codons and thus generate nonsense mutations. Perhaps that is just as well as the protein would probably be too garbled anyway to be useful to the cell.

Indels of three nucleotides or multiples of three may be less serious because they preserve the reading frame (see **Patient E** above).

However, a number of inherited human disorders are caused by the insertion of many copies of the same triplet of nucleotides. **Huntington's disease** and the **fragile X syndrome** are examples of such **trinucleotide repeat** diseases.

Fragile X Syndrome

Several disorders in humans are caused by the inheritance of genes that have undergone insertions of a string of 3 or 4 nucleotides repeated over and over. A locus on the human X chromosome contains such a stretch of nucleotides in which the triplet **CGG** is repeated (CGGCGGCGGCGG, etc.). The number of CGGs may be as few as 5 or as many as 50 without causing a harmful phenotype (these repeated nucleotides are in a noncoding region of the gene). Even 100 repeats usually cause no harm. However, these longer repeats have a tendency to grow longer still from one generation to the next (to as many as 4000 repeats).



This causes a constriction in the X chromosome, which makes it quite fragile. Males who inherit such a chromosome (only from their mothers, of course) show a number of harmful phenotypic effects including mental retardation. Females who inherit a fragile X (also from their mothers; males with the syndrome seldom become fathers) are only mildly affected.

This image shows the pattern of inheritance of the fragile X syndrome in one family. The number of times that the trinucleotide **CGG** is repeated is given under the symbols. The gene is on the X chromosome, so women (circles) have two copies of it; men (squares) have only one. People with a gene containing 80–90 repeats are normal (light red), but this gene is unstable, and the number of repeats can increase into the hundreds in their offspring. Males who inherit such an enlarged gene suffer from the syndrome (solid red squares). (Data from C. T. Caskey, et al.).

Polyglutamine Diseases

In these disorders, the repeated trinucleotide is **CAG**, which adds a string of glutamines (Gln) to the encoded protein. These have beeen implicated in a number of central nervous system disorders including

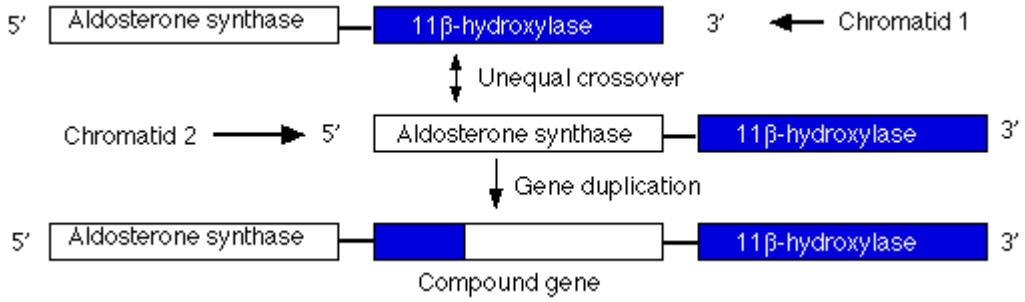
- **Huntington's disease** (where the protein called **huntingtin** carries the extra glutamines). The abnormal protein increases the level of the p53 protein in brain cells causing their death by apoptosis.
- some cases of Parkinson's disease where the extra glutamines are in the protein ataxin-2;
- some case of amyotrophic lateral sclerosis (ALS) — again where ataxin-2 is the culprit. (ALS is often called "Lou Gehrig's disease" after the baseball player who died from it.)

Muscular Dystrophy

Some forms of muscular dystrophy that appear in adults are caused by tri- or tetranucleotide, e.g. $(CTG)_n$ and $(CCTG)_n$, repeats where n may run into the thousands. The huge RNA transcripts that result interfere with the alternative splicing of other transcripts in the nucleus.

Duplications

Duplications are a doubling of a section of the genome. During meiosis, crossing over between sister chromatids that are out of alignment can produce one chromatid with an duplicated gene and the other (not shown) having two genes with deletions. In the case shown here, unequal crossing over created a second copy of a gene needed for the synthesis of the steroid hormone **aldosterone**.



However, this new gene carries inappropriate promoters at its 5' end (acquired from the 11-beta hydroxylase gene) that cause it to be expressed more strongly than the normal gene. The mutant gene is dominant: all members of one family (through four generations) who inherited at least one chromosome carrying this duplication suffered from high blood pressure and were prone to early death from stroke.

Gene duplication has also been implicated in several human neurological disorders.

Gene duplication has occurred repeatedly during the evolution of eukaryotes. Genome analysis reveals many genes with similar sequences in a single organism. Presumably these paralogous genes have arisen by repeated duplication of an ancestral gene.

Such gene duplication can be beneficial.

- Over time, the duplicates can acquire different functions.
- The proteins they encode can take on different functions; for example, if the original gene product carried out two different functions (see "pleiotropy"), each duplicated gene can now specialize at one function and do a better job at it than the parental gene.
- But even if they do not, changes in the regulatory sequences of the genes (promoters and enhancers) may cause the same protein to be expressed at different times and/or in different tissues.

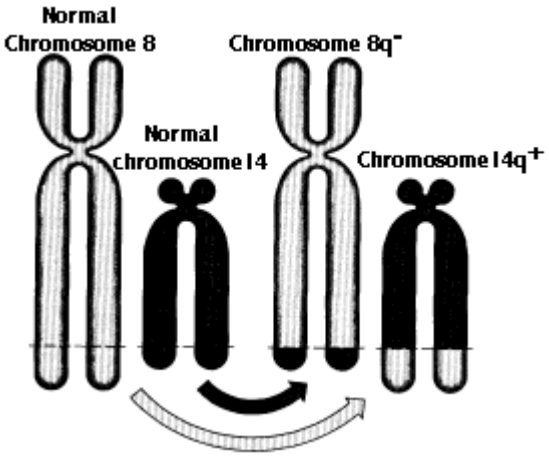
Either situation can provide the basis for adaptive evolution.

- But even while two paralogous genes are still similar in sequence and function, their existence provides **redundancy** ("belt and suspenders"). This may be a major reason why knocking out genes in yeast, "knockout mice", etc. so often has such a mild effect on the phenotype. The function of the knocked out gene can be taken over by a paralog.
- After gene duplication, random loss — or inactivation — of one of these genes at a later time in
- one group of descendants
- different from the loss in another group

could provide a barrier (a "post-zygotic isolating mechanism") to the two groups interbreeding. Such a barrier could cause speciation: the evolution of two different species from a single ancestral species.

Translocations

Translocations are the transfer of a piece of one chromosome to a **nonhomologous chromosome**. Translocations are often reciprocal; that is, the two nonhomologues swap segments.



Translocations can alter the phenotype is several ways:

- the break may occur **within** a gene destroying its function
- translocated genes may come under the influence of different promoters and enhancers so that their expression is altered. The translocations in Burkitt's lymphoma are an example.
- the breakpoint may occur **within a gene** creating a hybrid gene. This may be transcribed and translated into a protein with an N-terminal of one normal cell protein coupled to the C-terminal of another. The Philadelphia chromosome found so often in the leukemic cells of patients with **chronic myelogenous leukemia** (**CML**) is the result of a translocation which produces a compound gene (**bcr-abl**).

Frequency of Mutations

Mutations are rare events.

This is surprising. Humans inherit 3 x 10^9 base pairs of DNA from each parent. Just considering single-base substitutions, this means that each cell has 6 billion (6 x 10^9) different base pairs that can be the target of a substitution.

Single-base substitutions are most apt to occur when DNA is being copied; for eukaryotes that means during S phase of the cell cycle.

No process is 100% accurate. Even the most highly skilled typist will introduce errors when copying a manuscript. So it is with DNA replication. Like a conscientious typist, the cell does proofread the accuracy of its copy. But, even so, errors slip through.

It has been estimated that in humans and other mammals, uncorrected errors (= mutations) occur at the rate of about 1 in every 50 million (5 x 10^7) nucleotides added to the chain. (Not bad — I wish that I could type so accurately.) But with 6 x 10^9 base pairs in a human cell, that mean that each new cell contains some 120 new mutations.

Should we be worried? Probably not.

Most (as much as 97%) of our DNA does not encode anything. This includes:

- repetitive DNA like Alu elements and other so-called "junk" DNA
- noncoding DNA in introns and flanking structural genes. (However, mutations here **can** have an effect by altering the expression of the gene or interfering with correct splicing of the gene's mRNA.)
- Even in coding regions, the existence of synonymous codons may result in the altered (mutated) gene still encoding the same amino acid in the protein.

How can we measure the frequency at which phenotype-altering mutations occur? In humans, it is not easy.

- First we must be sure that the mutation is **newly-arisen**. (Some populations have high frequencies of a particular mutation, not because the gene is especially susceptible, but because it has been passed down through the generations from a early "founder". [Link to an example]).
- Recessive mutations (most of them are) will not be seen except on the rare occasions that both parents contribute a mutation at the same locus to their child.
- This leaves us with estimating mutation frequencies for genes that are inherited as
- **autosomal dominants**
- **X-linked recessives**; that is, recessives on the X chromosome which will be expressed in males because they inherit only one X chromosome.

Some Examples (expressed as the frequency of mutations occurring at that locus in the gametes)

- **Autosomal dominants**
- **Retinoblastoma**
  in the ***RB*** gene [Link]: about 8 per million (8×10^{-6})
- **Osteogenesis imperfecta**
  in one or the other of the two genes that encode Type I **collagen** [Link]: about 1 per 100,000 (10^{-5})
- Inherited tendency to polyps (and later cancer) in the colon.
  in a tumor suppressor gene (*APC*) [Link]: ~10^{-5}
- **X-linked recessives**
- Hemophilia A [Link]
  ~3×10^{-5} (the **Factor VIII** gene)
- Duchenne Muscular Dystrophy (DMD) [Link]
  >8×10^{-5} (the **dystrophin** gene)
  Why should the mutation frequency in the dystrophin gene be so much larger than most of the others? It's probably a matter of size. The dystrophin gene stretches over 2.3×10^6 base pairs of DNA. This is almost 0.1% of the entire human genome! Such a huge gene offers many possibilities for damage.

Males Contribute More Mutations Than Females

If most mutations occur during S phase of cell division, then males should be more at risk. This is because

> only two dozen (24) or so mitotic divisions occur from the fertilized egg that starts a little girl's embryonic development and the setting aside of her future eggs (which is done long before she is even born).
>
> The sperm of 30-year old man, in contrast, is the descendant of at least 400 mitotic divisions since the fertilized egg that formed him.

So,

- fathers are more likely than mothers to transmit newly-formed mutations to their children. (But **chromosomal aberrations**, like aneuploidy, are more apt to arise in eggs than in sperm.)
- The children of aged fathers suffer more genetic disorders than those of young fathers.
- Actual measurements show that this phenomenon of **"male bias"** is not as bad as the numbers suggest. Possible reasons:
- Perhaps many mutations (e.g., those caused by chemicals within the cell or by radiation) occur independently of DNA replication and thus would affect males and females equally.
- Even in an older man, fresh sperm may come from precursor stem cells that have been held in "reserve" and are not the result of years of mitotic divisions.
- Evolution may have led to mechanisms that enhance the accuracy of DNA repair in the precursors of sperm.

Somatic vs. Germline Mutations

The significance of mutations is profoundly influenced by the distinction between germline and soma. Mutations that occur in a **somatic cell**, in the bone marrow or liver for example, may

- damage the cell
- make the cell cancerous
- kill the cell

Whatever the effect, the ultimate fate of that **somatic mutation** is to disappear when the cell in which it occurred, or its owner, dies.

Germline mutations, in contrast, will be found in every cell descended from the zygote to which that mutant gamete contributed. If an adult is successfully produced, every one of its cells will contain the mutation. Included among these will be the next generation of gametes, so if the owner is able to become a parent, that mutation will pass down to yet another generation.