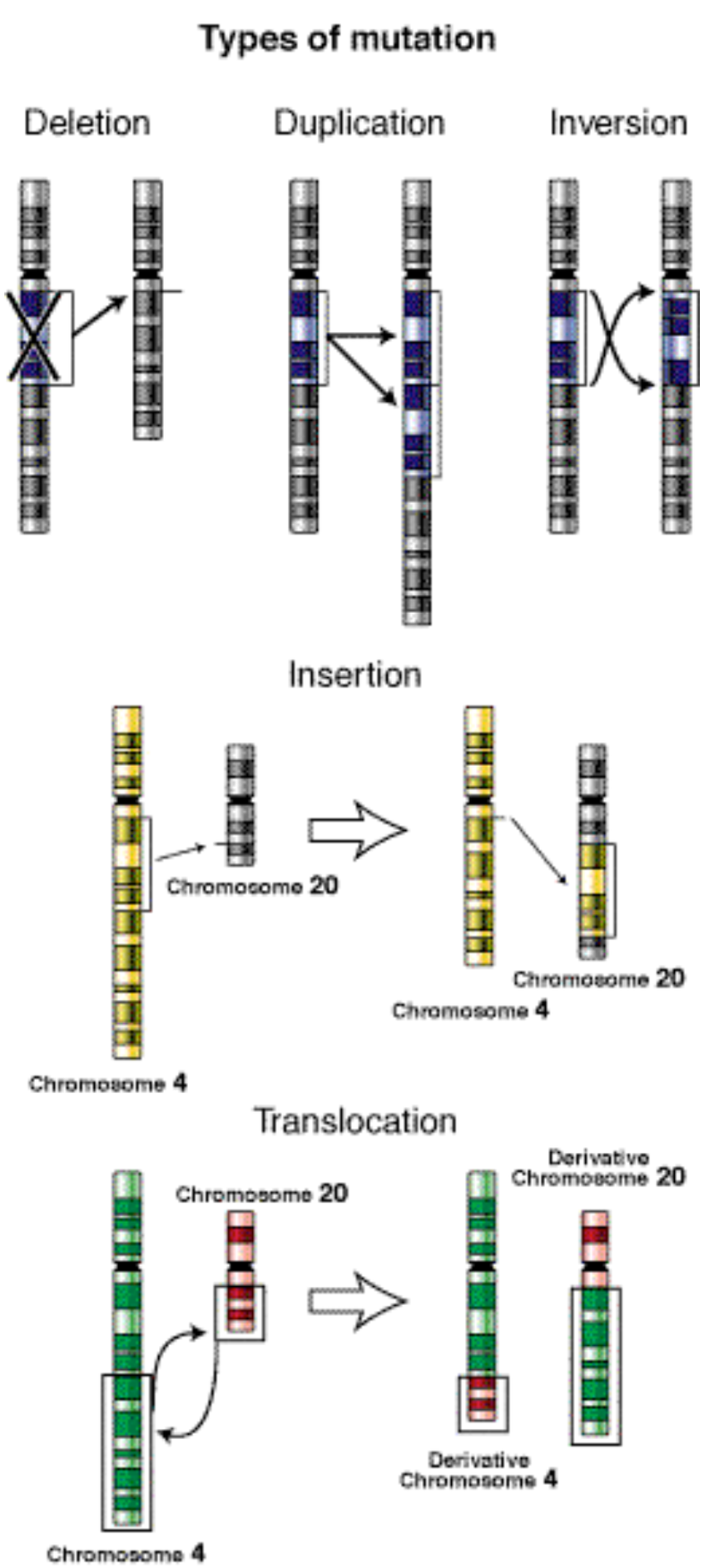
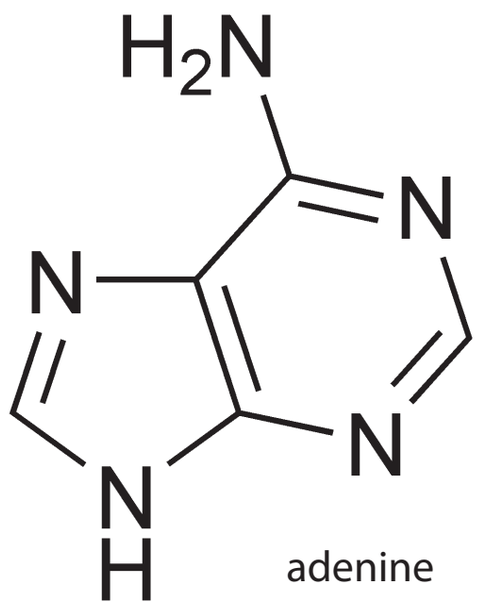
Type of mutations

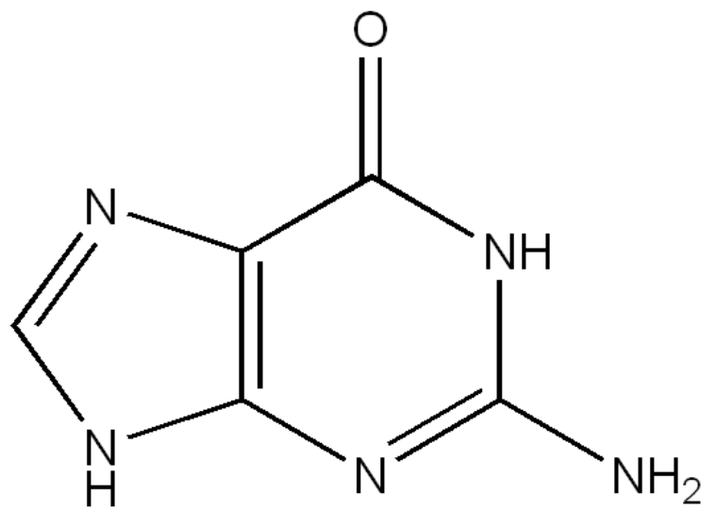
- Deletion
- Duplication
- Inversion
- Insertion
- Translocation
- Point mutations
  - ★ Silent mutations: does not change amino acid.
  - ★ Missense mutations: different amino acid.
  - ★ Nonsense mutations: code for a stop, truncate protein.



Purine

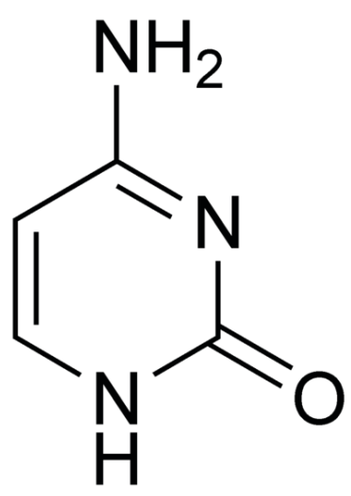


adenine

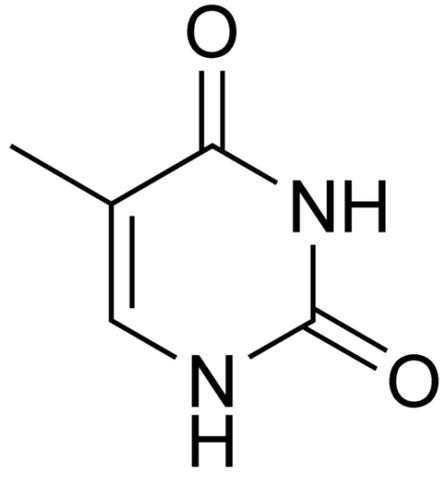


guanine

Pyrimidine



Cytosine



Thymine

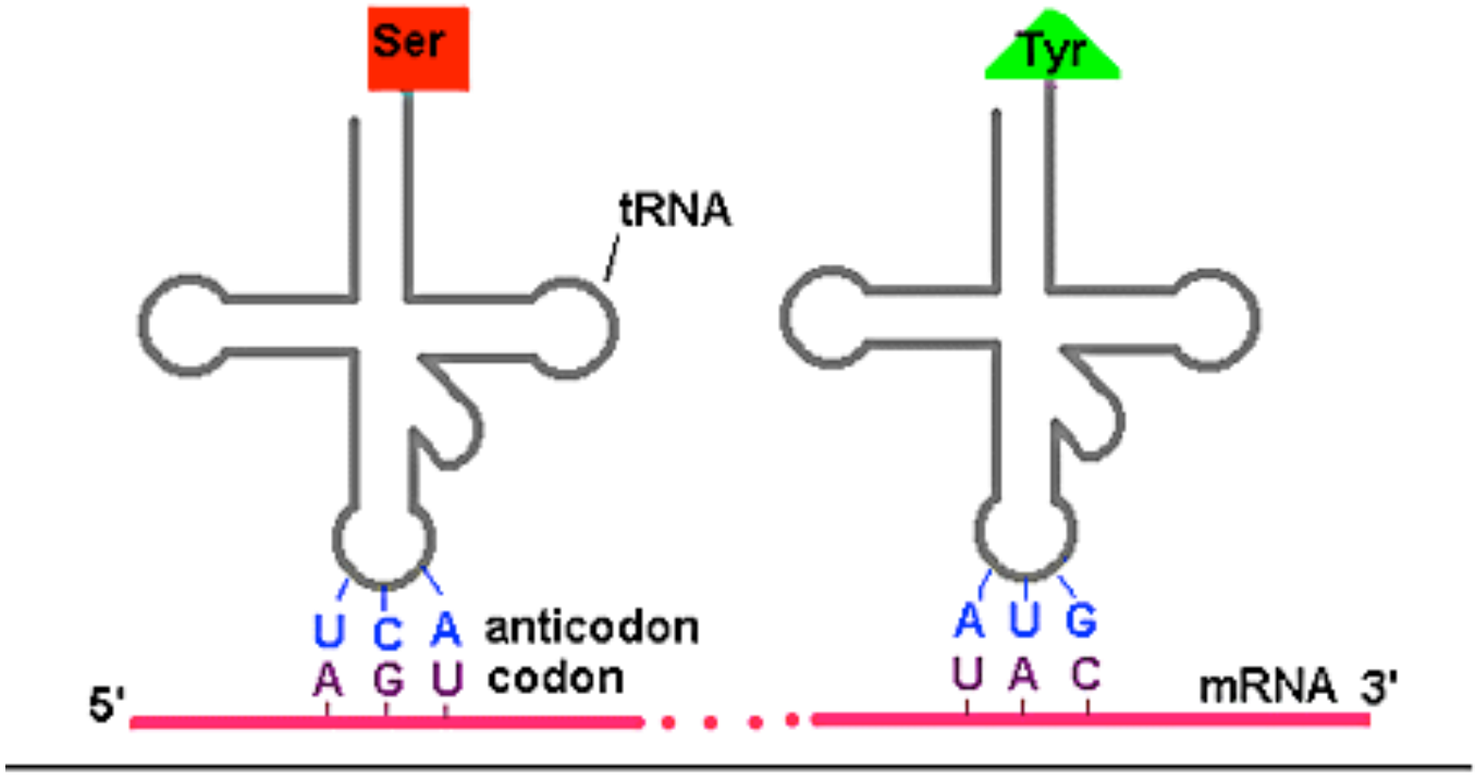
Examples of base-pair substitutions:

purine substituted with a different purine (A → G) or a pyrimidine for a different pyrimidine (C → T).

This type of substitution mutation is a transition.

Another kind is a transversion, which is less common.

In transversion, purine is substituted with pyrimidine or a pyrimidine with a purine.



1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The Genetic Code

U=uracil
in RNA
equivalent to
T=Thymine in
DNA

Wild type allele:

M D D Q S R M L Q T L A G V N L

atggacgatcaatccaggatgctgcagactctggccggggtgaacctg

silent (third base pair) mutation:

M D D Q S R M L Q T L A G V N L

atggacgatcaatccaggatgctgcaaactctggccggggtgaacctg

point mutation (missense):

M D D Q S R M L K T L A G V N L

atggacgatcaatccaggatgctgaagactctggccggggtgaacctg

point mutation (nonsense):

M D D Q S R M L stop

atggacgatcaatccaggatgctgtagactctggccggggtgaacctg

frameshift leading to premature termination:

M D D Q S R M L R L W P G stop

atggacgatcaatccaggatgctgagactctggccggggtgaacctg

c was excised

switch to handout

Scaling of branch lengths [edit]

By comparing extant sequences, one can determine the amount of sequence divergence. This raw measurement of divergence provides information about the number of changes that have occurred along the path separating the sequences. The simple count of differences (the Hamming distance) between sequences will often underestimate the number of substitution because of multiple hits (see homoplasy). Trying to estimate the exact number of changes that have occurred is difficult, and usually not necessary. Instead, branch lengths (and path lengths) in phylogenetic analyses are usually expressed in the expected number of changes per site. The path length is the product of the duration of the path in time and the mean rate of substitutions. While their product can be estimated, the rate and time are not identifiable from sequence divergence.

The descriptions of rate matrices on this page accurately reflect the relative magnitude of different substitutions, but these rate matrices are **not** scaled such that a branch length of 1 yields one expected change. This scaling can be accomplished by multiplying every element of the matrix by the same factor, or simply by scaling the branch lengths. If we use the β to denote the scaling factor, and v to denote the branch length measured in the expected number of substitutions per site then βv is used in the transition probability formulae below in place of μt. Note that v is a parameter to be estimated from data, and is referred to as the branch length, while β is simply a number that can be calculated from the rate matrix (it is not a separate free parameter).

The value of β can be found by forcing the expected rate of flux of states to 1. The diagonal entries of the rate-matrix (the Q matrix) represent -1 times the rate of leaving each state. For time-reversible models, we know the equilibrium state frequencies (these are simply the π_i parameter value for state i). Thus we can find the expected rate of change by calculating the sum of flux out of each state weighted by the proportion of sites that are expected to be in that class. Setting β to be the reciprocal of this sum will guarantee that scaled process has an expected flux of 1:

$$\beta = 1 / \left(-\sum_i \pi_i \mu_{ii} \right)$$

For example, in the Jukes-Cantor, the scaling factor would be $4/(3\mu)$ because the rate of leaving each state is $3\mu/4$.