Evolution lecture #5 -- **Molecular genetics and molecular evolution** -- Nov. 14th, 2007
(ch. 16: 296-307 (overview); ch. 17: 309-326 (overview); ch. 25: 499-508)

I. Summary of topics to be covered:

- Review the key features of DNA structure and the processes of gene transcription and translation (DNA of genes to amino acids of proteins)
- **Molecular features** of all life forms that support evolution
- Describe why phylogenetic trees drawn from molecular data should show the same broad patterns as those drawn from fossil data
- Use the molecular clock principle to estimate divergence times between groups on phylogenetic trees

II. Molecular genetics

**DNA:** deoxyribonucleic acid; the genetic material.

**nucleotides:** adenine (A), cytosine (C), guanine (G), thymine (T); in the DNA double helix, A pairs with T, and G pairs with C (see Fig. 16.7 (7th) (Fig. 16.5 6th))

**RNA:** ribonucleic acid; uracil (U) replaces thymine (T).

**mRNA:** messenger RNA; kind of RNA produced by transcription from the DNA and which acts as the message that is decoded to form proteins (Fig. 17.4 (7th) (Fig. 17.3 6th)).

**tRNA:** transfer RNA; kind of RNA that brings the amino acids to the ribosomes to make proteins. A transfer RNA molecule has an amino acid attached to it, and has attached in a different area of the molecule the anti-codon corresponding to that amino acid. In protein synthesis, each codon in the mRNA combines with the appropriate tRNA's anti-codon, and the amino acids are thus arranged in order and make the protein (Figs. 17.13 - 17.16 (7th) (Figs. 17.12 - 17.15 (6th))).

**amino acids:** the unit building blocks of proteins; a protein is a chain of amino acids in a certain sequence. There are 20 amino acids in the proteins of living things.

**genetic code:** the code relating the nucleotide **triplets** in the messenger RNA to amino acids in the protein (Fig. 17.5 (7th) (Fig. 17.4 6th)).

**codon:** a triplet of bases (nucleotides) in RNA coding for one amino acid. Conventionally, the triplet in the mRNA is the codon, and the triplet in the tRNA is the anti-codon.
**degenerate:** more than one codon can code for one amino acid; hence some mutations do not result in an amino acid change; these are called synonymous mutations, as distinct from nonsynonymous mutations which do result in an amino acid change.

**universal:** the same genetic code is used in all organisms (with a few small exceptions).

**Exons vs. Introns,** and alternative splicing. Some genes can encode more than one kind of polypeptide, depending on which segments are treated as exons during RNA splicing. Such variations are called alternative RNA splicing. Because of alternative splicing, the number of different proteins an organism can produce is much greater than its number of genes.

**mutations:** a heritable change in DNA. Mutations can be deleterious, neutral, or selected. Mutations can involve a change of a nucleotide (base change) as well as insertions or deletions of genetic material (see Fig. 25.6 (7th)).

### III. Molecular features that support evolution

- Common molecular and biochemical features of all life forms
- Universal use of DNA as the genetic material, rules of genetic transmission, and genetic code
- Universal processes of gene expression, protein synthesis, and protein function
- The genes and their functions are strikingly similar in different organisms
- DNA shows evidence of variation (diversity), the continuity of life, and the unitary origin of life
- The more closely related two species are to each other, the more similar their DNA, and vice versa (humans and chimps are more closely related to each other than either is to gorillas or orangutans) (Fig. 34.38 (7th) (Fig. 34.35 (6th)))
- Evolutionary trees based on DNA are strikingly similar to those based on anatomical, developmental, and fossil evidence

### IV. Genome sequences

The complete DNA genome sequences of more than 300 organisms have been completed, mostly prokaryotes, but many eukaryotes as well, for example, yeast, fruit fly, mustard plant, moss, rice, nematode worm, mouse, and humans, and sequencing of the genomes of thousands of organisms is in progress (http://www.ncbi.nlm.nih.gov/Genomes/index.html).
Mitochondria and chloroplast have their own genomes, in addition to the nuclear genome. These are descended from the endosymbiotic bacteria that originally came to live inside the eukaryotic cell. A large number of these have been sequenced, and provide important phylogentic characters, both in their nucleotide sequence and in their gene order. Below is an example of the latter from green plants, to be explained in lecture.

### Change in Size of Inverted Repeat

<table>
<thead>
<tr>
<th>Species</th>
<th>rpl19</th>
<th>rpl2 rpl23trnl</th>
<th>trnVrrn16 trnI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chaetosphaeridium</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>trnVrrn16 trnI</td>
</tr>
<tr>
<td>Marchantia</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>trnVrrn16 trnI</td>
</tr>
<tr>
<td>Physcomitrella</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>trnVrrn16 trnI</td>
</tr>
<tr>
<td>Anthoceros</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>ndhB rps7 rps12 trnVrrn16 trnI</td>
</tr>
<tr>
<td>Huperzia</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>rps12 trnVrrn16 trnI</td>
</tr>
<tr>
<td>Isoetes</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>rps7 rps12 trnVrrn16 trnI</td>
</tr>
<tr>
<td>Angiopteris</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>trnL ndhB rps7 rps12 trnVrrn16 trnI</td>
</tr>
<tr>
<td>Nicotiana</td>
<td>rpl19</td>
<td>rpl2 rpl23trnl</td>
<td>ycf2 ycf15 trnL ndhB rps7 rps12 trnVrrn16 trnI</td>
</tr>
<tr>
<td>Zea</td>
<td>rps19 trnH rpl2 rpl23trnl</td>
<td>trnL ndhB rps7 trnVrrn16 trnI</td>
<td></td>
</tr>
</tbody>
</table>

**Character numbers:** 25 10 26 27 28 43 54 1

### V. Molecular evolution

**evolutionary trees from DNA data:** A series of evolutionary changes involves a **progressive accumulation** of genetic change in the DNA.

Even if there were no selection operating, i.e., mutations which arose were selectively neutral, because of the finite size of populations and consequent chance events, alleles always eventually become lost from a population, so there is eventual replacement of allelic types by another. (This will be covered later in section on genetic drift.)

The more distantly related two species are the more genetic differences (amino acid changes or nucleotide changes) that will have accumulated between them. So, the longer the time since the organisms diverged, the greater the number of differences in the nucleotide sequence of the gene, e.g., cytochrome c.
The gene that encodes small subunit ribosomal RNA (SSU rRNA) was extensively used for the classification of microorganisms and resulted in recognition of archae and bacteria as two distinct domains within the prokaryotes.

Evolutionary trees drawn from DNA data agree well with those drawn from the fossil record, and can be important where convergent evolution of similar characteristics can cause confusion in drawing evolutionary trees based on the characteristics of organisms, and/or when the fossil record is poor.

Almost any type of character (for example, morphological structures, characteristics of cells, biochemical pathways, genes, amino acids or nucleotides) can be used for inferring phylogenies, provided that they are homologous.

In addition to direct comparison of specific genes shared by different species, complex characters known are rare genomic changes (RGCs), that have a very low probability of being the result of convergence, can also be analyzed. As well as gene order, such RGCs include intron positions, insertions and deletions of genetic material (indels), retroposon (SINE and LINE) integrations, and gene fusion and fission events.

**substitution times:** the time required for substitution of different amino acids is long, in the millions of years. For example, with hemoglobin we expect on average 1 amino acid substitution every 5 million years.

**molecular clock:** there is a regularity with time in the rate of accumulation of genetic changes at the nucleotide and amino acid levels (Fig. 25.17 (6th)). For example, see this distance matrix (i.e., number of amino acid difference in some particular protein):

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Gorilla</th>
<th>Pig</th>
<th>Rabbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>Gorilla</td>
<td>0</td>
<td>20</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>0</td>
<td></td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Rabbit</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Different proteins show different rates of accumulation of changes. Some evolve very slowly, such as histones, others at a medium rate such as hemoglobin, and others faster such as fibrinopeptide. But for each protein, the molecular clock roughly applies.

Non-functional DNA, including pseudogenes, is expected to evolve much more quickly than coding region DNA, where many mutations will be deleterious.

**divergence time between two groups:** when the divergence time between two groups is unknown, the molecular clock principle can be applied to obtain an estimate of their divergence times.
The first estimate of time to the most recent common ancestor of humans and chimps was obtained using molecular data (~ 5-6 Myr); new hominid fossil discoveries since then are in agreement with the molecular data.

The molecular clock is not constant; different lineages can show faster, or slower rates of evolution than other lineages. Trees with different branch lengths (phylogenograms (Fig. 25.12 (7th)) document these different rates; with ultrametric trees (Fig. 25.13 (7th)) all branches have the same total length from bottom of the tree to the twigs.

**synonymous and nonsynonymous mutations:** comparison of the rates of accumulation of synonymous (no amino acid change) and nonsynonymous (results in an amino acid change) changes can indicate selection.

**codon usage bias:** even though certain codons are theoretically synonymous, we often find that organisms discriminate among them.

**genome evolution:** various mechanisms such as duplication of genetic material followed by diversification and selection, have played a role in the evolution of complex genomes:

1. exons duplicate or shuffle to change the size or function of genes (the coding regions of genes are divided into separate regions called exons, separated by regions called introns (which are spliced out of the messenger RNA before translation)),
2. entire genes duplicate to create multigene families,
3. multigene families duplicate to produce multigene superfamilies, and
4. the entire genome duplicates to double the number and copies of every gene and gene family (called "polyploidy", can be "autopolyploid" or "allopolyploid").

**orthologous genes:** refer to homologous genes found in different species, for example, the cytochrome c gene. Orthologous genes are widespread and can extend over huge evolutionary distances. In mice and humans about 99% of the genes are orthologous, and 50% of human genes are orthologous with those of yeast.

**paralogous genes:** result from gene duplication, for example olfactory receptor genes. This duplication often leads to specialization of the two copies within the organisms, much like speciation can lead to diversification of ecosystems. Genes and “species,” or other higher-order lineages thus may have different histories.
**Transition-transversion bias**: results from the physical chemistry of the nucleotide bases -- substitutions are easier within category:

- purines: A G
- pyrimidines: C T U

**The universal tree of life**: see Fig. 25.18 (7th) (Figs. 28.6 and 28.7 (6th)). The early history of life of the bacteria and archae is not clear; it appears that there was substantial interchange of genes between organisms (horizontal gene transfer).

**Questions relating to lecture on molecular genetics and molecular evolution**