**MOLECULAR EVOLUTION**

**Problem:** We cannot observe evolution at work because processes are too slow. Therefore, have to rely on interpretation of long past events. How can this be done?

We need to identify **homologous** characters in organisms we want to compare. Homology means similarity of characters by common descent. An example of such characters is the numbers of bones and their arrangement relative to one another in vertebrate limbs (see Figure 21.2, p. 415). Once we have identified these characters in all organisms of interest we can compare these structures and order them according to their similarity.

Morphological characters have been used extensively to reconstruct phylogenetic relationships but this approach only works if there is a well-preserved fossil record, and the organisms are neither too closely nor too distantly related. Particularly, for microorganisms, which are unicellular and do not have a good fossil record, morphological criteria cannot be used to determine phylogenetic relationships (see p. 485, “Morphological Diversity”).

However, genes can be regarded as **molecular chronometers** (see also p. 470, Box 24.1). Since they accumulate point mutations over time their sequence difference is roughly proportional to evolutionary distance between the organisms examined.

Genes for this type of analysis must:

- be present in all organisms examined
- have a common ancestry
- have a conserved function
- be a mosaic of functionally constraint stretches of nucleotides and stretches where alterations are “allowed” – this helps in identifying homologous nucleotide positions in different organisms

Examples for genes that are present in ALL organisms and that have been used extensively in reconstruction of phylogeny are the 16S ribosomal RNAs (rRNAs) (see also p. 486, “Molecular Phylogenies”). The conserved function of the rRNAs is reflected in their secondary structure. The molecules play a structural role in the ribosome in all organisms. In order to fulfill their function, they must fold into a certain configuration in which three types of sequence regions can be identified: (A) constrained to an extent that no sequence change is possible, (B) constrained in length and morphology but sequence change is possible, and (C) so minimally constrained that both variation in sequence and length is possible. It is the regions of the A and B type that are most useful for phylogenetic analysis.
**Figure 2.** Two secondary structure models of bacterial and archaeal 16S rRNA. Note that the overall ‘morphology’ of the molecule is very similar despite the fact that the last common ancestor of these organisms dates back several billion years. The lettering denotes examples of sequence stretches as explained above.

**How are genes used to determine phylogeny?**

Imagine a simple stretch of sequence that occurs in three different species. You can order the sequences by similarity by simple inspection and then connect them with lines connecting more similar sequences with shorter lines. This is a simple phylogenetic tree, which has the underlying assumption that the more similar sequences are the less time has passed since they had a common ancestor.

```
<table>
<thead>
<tr>
<th></th>
<th>species 1</th>
<th></th>
<th>species 2</th>
<th></th>
<th>species 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATCGAA</td>
<td></td>
<td>AACGAA</td>
<td></td>
<td>ATGCTT</td>
</tr>
</tbody>
</table>
```

"Phylogenetic Tree"

But there are several complicating factors.

- genes are more complex (longer)
- there is length variation due to insertions and deletions
- single positions can be hit by multiple mutational events
Need computer and use complicated algorithms to figure out relationships. There are three basic steps in determining the molecular phylogeny of organisms.

• 1st step is **sequence alignment**: homologous bases are lined up in columns to facilitate comparison.

• 2nd step is the determination of a pairwise **evolutionary distance** by counting the total number of homologous bases in the alignment and by calculating the fraction of sequence difference for each sequence pair.
Figure 3. Steps involved in phylogenetic tree construction. See above for explanation.
• 3rd step is the statistical **correction** of the evolutionary distance for possibilities of multiple mutational events in the sequences.

• 4th step is the drawing of a **'best-fit' tree** that reflects the sequence distance between the organisms examined.

These steps are illustrated in the figures below. Figure 3 illustrates the individual steps in the phylogenetic distance analysis. Figure 4 shows a realistic alignment.

---

**Figure 4.** Part of an alignment of ribosomal rRNA genes used to determine evolutionary relationships.

---

**The tree of life**

Molecular phylogeny has fundamentally changed the way we look at the evolution of all living organisms. The classical view of evolution of all living organisms held that there were either two or five kingdoms based on morphological criteria.
Carl Wose used 16S ribosomal RNA (rRNA) sequences and found that bacteria really consist of two groups, which are ± equidistantly related to eukaryotes.
Figure 5 shows a ‘universal’ tree of life where it becomes apparent that life evolved into three groups two of which are prokaryotic. Thus, multicellular organisms are in evolutionary terms a very recent development. This also entails that the highest the diversity of organisms is found among
microorganisms. Indeed, the molecular approach to diversity has revolutionized the way we look at microorganisms in the environment.

How the 16S rRNA approach has helped us to understand the diversity of bacteria in the environment:

**Problem:** Bacteria are small and all look very similar - cannot be identified by just looking at them under the microscope. Therefore, bacteria had to be isolated in pure culture and their physiology and metabolism studied to be able to identify them.

**BUT:** <1% of bacteria in nature grow easily in any given culture medium

**Solution:** Use genes to estimate the diversity
Result

- Poor correspondence between culture collections and organisms identified by environmental DNA sequences.
- Bacterial diversity has been greatly underestimated, wherever we look we find previously uncultured microorganisms, e.g., 1 g of forest soil 10,000 bacterial species