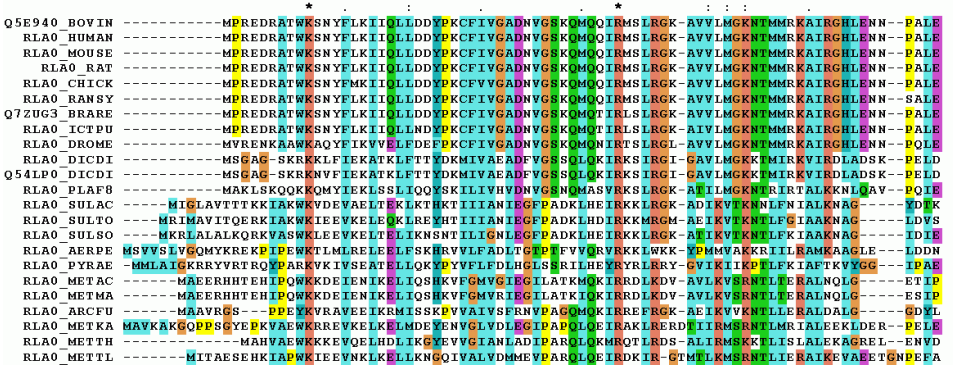
11. Multiple Sequence Alignment

Outline

Multiple sequence alignment

- multiple sequence alignment (MSA) is sequence alignment of three or more biological sequences such as DNA, RNA, or protein
- an example protein MSA



Multiple sequence alignment

- each row of the MSA corresponds to the sequence of a specific protein
- each column of the MSA corresponds to a position in the sequence
- dash symbol means the sequence does not have an amino acid aligned at that position
- protein sequences are in the same MSA are evolutionarily related: they are homologous
- homologous sequences are derived from a common ancestor, so they are similar in sequence, structure, and function

Multiple sequence alignment

- MSA of a protein contains more information than a single sequence
- can be used to identify conserved regions in the protein
- conserved regions are often important for the protein's function
- used to infer the evolutionary relationships between the sequences
- used to search for homologous sequences in a database
- used to predict the structure and function of a protein

Multiple sequence alignment

- multiple algorithms exist for constructing MSAs
- most algorithms require a query sequence and a database of sequences
- they iteratively search for homologous sequences in the database and align them
- example algorithms: Clustal Omega, MUSCLE

Protein family

- a protein family is a group of proteins that share a common evolutionary origin
- members of a protein family are homologous and have similar sequences, structures, and functions
- sequences of a protein family are aligned to create a multiple sequence alignment
- the Pfam database is a collection of protein families

Outline