- We are given a set of sequences:
- Problem: Which sequences are related? How sure can I be?
- Homology detection

- We are given a set of related sequences:
- Problem: Which positions belong together? How sure can I be?
- Alignment

Mutations and mismatch positions

- Mutations have a direction (a ’G’ changes into an ’A’)
- Mutations correspond to mismatch positions in an alignment
- Mismatch positions have no direction
- Mismatch positions are the result of an assignment process, done by a human being or a computer
- Mismatch positions are inferred from sequence similarities in there neighborhood

Deletions insertions and gaps

- Deletions insertions have a direction
- Both deletions and insertions are represented by gaps in an alignment
- Gaps have no direction
- Gaps are not part of a sequence, they are introduced.
- Gaps are inferred from sequence similarities in there neighborhood

- We only observe sequences from organisms that live today
- We do not observe mutations deletions or insertions
- We only observe sequence similarities that are the result of a common ancestor sequence and a series of mutations deletions or insertions
- Our aim is to infer the evolutionary history of the sequences and to learn from conservation patterns

SEQUENCE ANALYSIS IS STATISTICS

- Sequence are strings over a finite alphabet
- Sequence comparison is too complex to be done manually
- It is done by string algorithms

SEQUENCE ANALYSIS IS COMPUTATION

- Biological sequence encode for metabolic functions
- The fitness of an organism is determined by the performance of its metabolism
- Selection is determined by fitness
- Selection gives 'shape' to the sequences

SEQUENCE ANALYSIS IS BIOLOGY