



Introduction to Bioinformatics online course: IBT

Practical Assignment

Module name: Sequence alignment theory and applications

Session name: Introduction to sequence alignment

Trainer: Jonathan Kayondo, Sonal Henson

Participant: <write your name here>

Date: <write today's date here>

Introduction to Sequence Alignment

Introduction

Sequence alignment is a process that enables the comparative analysis of similarities and differences at nucleotide (gene) or amino acid (protein) level with the aim of inferring structural, functional and **evolutionary** relationships between/among sequences being studied. Alignments are mostly made with aid of computer programs using algorithms that assess quality of an alignment using a **scoring scheme** that rewards matches and punishes mismatches and gaps.

In the following exercise we practice, resolving various evolutionary relationships, and sequence alignment scoring using various reward/penalty rules.

Tools used in this session

- Lecture slides

Please note

- **Hand-in information** If you are formally enrolled in the IBT course, please upload your completed assignment to the Vula 'Assignments' tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

Task 1: Local and Global sequence alignment.

Task 1: instructions



Introduction to Bioinformatics online course: IBT

List two (2) major differences between a local and global sequence alignment strategy?

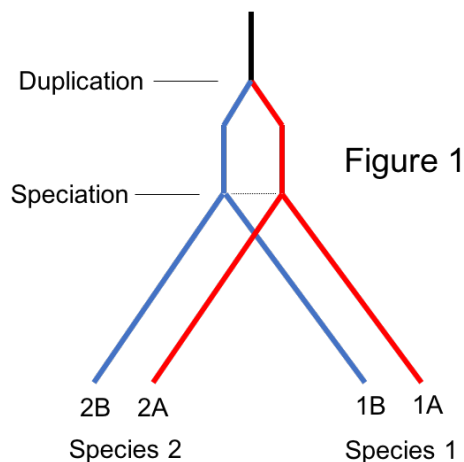
Task 1: participant's answer

<start typing your answer here>

Task 2: Homology – Orthology vs Paralogy

Task 2: instructions

Figure 1 shows the ancestry of a gene (black line), the last common ancestor of species 1 and species 2. The gene undergoes a gene duplication event, and both copies (A and B in red and blue, respectively), while gradually evolving apart, persist in the last common ancestor until a subsequent speciation event splits the lineage into two new species. Both species inherit both copies of the duplicated gene, where they continue to diverge until the present day.

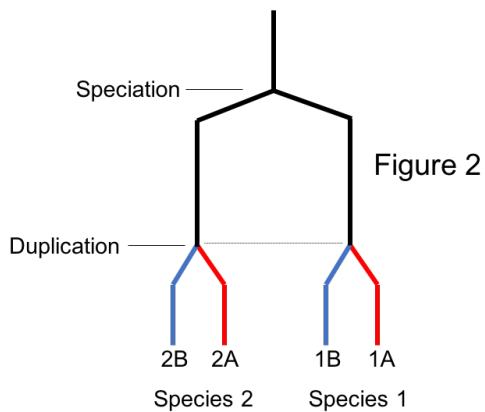


1. Are genes 1A and 1B orthologs or paralogs?
2. Are genes 1A and 2A orthologs or paralogs?
3. Are genes 1B and 2A orthologs or paralogs?

Similar to Figure 1, in figure 2 the ancestral gene (black line) first splits by speciation into two new species, both of which inherit the gene. Then independent gene duplication events create copies of the gene in each species.



Introduction to Bioinformatics online course: IBT



1. Are genes 1A and 1B orthologs or paralogs?
2. Are genes 1A and 2A orthologs or paralogs?
3. Are genes 1B and 2A orthologs or paralogs?

Task 2: participant's answer

<start typing your answer here>

Task 3: Understanding gap penalties –constant, linear and affine.

Task 3: instructions

Work out the total scores for the pair-wise alignments below given the scoring rules specified.

1. Alignment 1: Score rules: Match=+1, Mismatch = -1, Constant Gap= -2

```
1 TGGACGGTAATA 12
  |.||| .|||
1 TCGAC---CATA 9
```



Introduction to Bioinformatics online course: IBT

2. Alignment 2: Score rules: Match=+1, Mismatch = -1, Linear Gap= -1

```
1 CGGACTTGACT 12
  .||  ||.|| |
1 AGG--TTCTA-T 9
```

3. Alignment 3: Match = +5, Mismatch = -4, Gap open penalty = -7, Gap extension penalty = -2

```
1 CCTGCAATG---TAA 12
  |||  |||.  ||.
1 CCT--AATCTGCTAT 13
```

Task 3: participant's answer

<start typing your answer here>