

BIOINFORMATICS: GENOMES AND ALGORITHMS

Computer analysis of genetic information

François
Rechenmann



GENOMES AND ALGORITHMS

1. Genomic texts
- 2. Genes and proteins**
3. Gene prediction
4. Sequence comparison
5. Phylogenetic trees

2. Genes and proteins

- **The sequence as a model of DNA**
- Genes: from Mendel to molecular biology
- The genetic code
- A translation algorithm
- Implementing the genetic code
- Algorithms + data structures = programs
- The algorithm design trade-off
- DNA sequencing
- Whole genome sequencing
- How to find genes?

The sequence as a model of DNA

DNA is a molecule

- Organized in chromosomes, plasmids, segments,...
- Compacted within the cell
 - In the nucleus of eukaryotic cells
 - Directly in the cytoplasm in prokaryotic cells
- May be altered through interactions with other molecules
 - Example: methylation
- Is a sequence of characters

a good model of a DNA molecule?

What is a good model?

- An exact and complete model?
 - No model is an exact representation of its object
 - The object itself is the only exact model
 - Remember the example of the mythical map at scale 1:1
 - ✓ Unfeasible
 - ✓ Useless
- A model is an oriented simplified description of its object
- Oriented by the set of questions it is expected to answer

What is a good model?

- An exact and complete model?
 - No model is an exact representation of its object
 - The object itself is the only exact model
 - Remember the example of the mythical map at scale 1:1
 - ✓ Unfeasible
 - ✓ Useless
- A model is an oriented simplified description of its object
- Oriented by the set of questions it is expected to answer

A good model is a useful model

Is the sequence a useful model of DNA?

- It is a linear text which ignores the 3D structure of DNA
 - Distant regions in the text may be close in the molecule
- It retains the succession of nucleotides as the only information coding

Is the sequence a useful model of DNA?

- It is a linear text which ignores the 3D structure of DNA
 - Distant regions in the text may be close in the molecule
- It retains the succession of nucleotides as the only information coding

But, yes it is useful!
Very useful indeed.

Is the sequence a useful model of DNA?

- It is a linear text which ignores the 3D structure of DNA
 - Distant regions in the text may be close in the molecule
- It retains the succession of nucleotides as the only information coding

But, yes it is useful!
Very useful indeed.

- A previous example: the prediction of the origin of replication
- A new example to come: the prediction of genes,
or protein coding regions

Be careful with the vocabulary!

- The biologist: “DNA molecule”, “nucleotides”
- The bioinformatician: “sequence”, “bases”
- The computer scientist: “string”, “characters”