1. Genomic texts

- The cell, atom of the living world
- At the heart of the cell: the DNA macromolecule
- DNA codes for genetic information
- **What is an algorithm?**
- Counting nucleotides
- GC and AT contents of DNA sequence
- DNA walk
- Compressing the DNA walk
- Predicting the origin of DNA replication?
- Overlapping sliding window

François Rechenmann

What is an algorithm?

- A series of operations to be executed for solving a problem
  - Here, the computation of nucleotide frequencies in a genetic sequence
- Much more formal than a recipee

What is an algorithm?

- A series of operations to be executed for solving a problem
  - Here, the computation of nucleotide frequencies in a genetic sequence
- Much more formal than a recipee
- Expected properties of an algorithm
  - termination
  - relevence
  - efficiency

What is an algorithm?

- A series of operations to be executed for solving a problem
  - Here, the computation of nucleotide frequencies in a genetic sequence
- Much more formal than a recipee
- Expected properties of an algorithm
  - termination
  - relevence
  - efficiency
- If to be executed by a computer, the algorithm
         has to be written in a programming language
- In this course, written in pseudocode

What is an algorithm?

- A series of operations to be executed for solving a problem
  - Here, the computation of nucleotide frequencies in a genetic sequence
- Much more formal than a recipee
- Expected properties of an algorithm
  - termination
  - relevence
  - efficiency

There are numerous programming languages: Java, C++, Perl, Python, CAML,…

- If to be executed by a computer, the algorithm

    has to be written in a programming language
- In this course, written in pseudocode

Computing the nucleotide frequencies

- Compute
  - Number of A: nbA
  - Number of C: nbC
  - Number of G: nbG
  - Number of T: nbT
- Compute total number of letters, i.e. nucleotides: TotalNb
- Compute frequencies
  - nbA/TotalNb
  - nbC/TotalNb
  - nbG/TotalNb
  - nbT/TotalNb

The input of our algorithm

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC*