# 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?
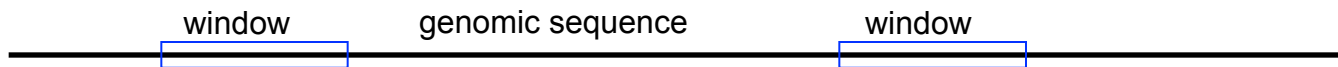- **Overlapping sliding window**

Overlapping sliding windows

Nucleotide frequencies along a sequence

- Compute nucleotide frequencies in a sliding window

window   genomic sequence

$$\frac{nbG - nbC}{nbG + nbC} = -0.3$$

window   genomic sequence   window
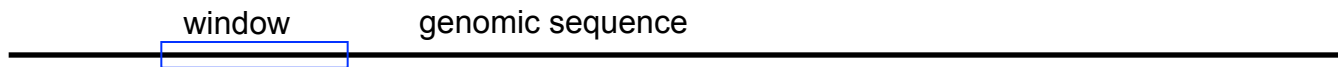
$$\frac{nbG - nbC}{nbG + nbC} = -0.3$$

$$\frac{nbG - nbC}{nbG + nbC} = +0.4$$

Variation of nucleotide frequencies along the genome

- Compute nucleotide frequencies in a sliding window
- Make the windows overlap

window

window

window

window

Variation of nucleotide frequencies along the genome

- Compute nucleotide ratio (G *vs.* C) in a sliding window
- Make the windows overlap

- Store the value of the ratio for each window
- Display the values as a curve,

    colinear with the sequence

```
SeqLength, L, J, InitW, nbC,nbG, OveLap: integer
sequence: character string [1:*]
RatioGC: array [1:*] of real


InitW, J ← 1


repeat
      nbC,nbG ← 0
      for i from InitW to min (InitW + L − 1, SeqLength) do
            case sequence [i] of
               "C": nbC ← nbC + 1
               "G": nbG ← nbG + 1
               "A", "T" :
            endcase
      endfor
      RatioGC [J] ← (nbG − nbC) / (nbG + nbC)
      J ← J + 1
      InitW ← InitW + (L - OverLap)
until InitW > SeqLength
```

window genomic sequence window

$$\frac{nbG - nbC}{nbG + nbC} = -0.3$$

$$\frac{nbG - nbC}{nbG + nbC} = +0.4$$

Value of the ratio

+1,0

0

450 000

900 000

Position on the sequence