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

What about the screen size?

- Resolution of a screen
  - The number of distinct pixels in each dimension that can be displayed
  - For example: 1024 x 768

- Problem:

How to fit a series of several millions or billions segments in one screen?

- Compression is the answer

```
SeqLength, L, InitW, nbA,nbC,nbG,nbT, NbStepsRight, NbStepsUp: integer
XEndSegment, YEndSegment, Step: real
sequence: character string [1:*]
nbA,nbC,nbG,nbT ← 0
InitW← 1
repeat
      for i from InitW to InitW + L – 1 do
           case sequence [i] of
              "A": nbA ← nbA + 1
              "C": nbC ← nbC + 1
              "G": nbG ← nbG + 1
              "T": nbT ← nbT + 1
           endcase
      endfor
      NbStepsRight ← nbC — NbG
      NbStepsUp ← nbA — nbT
      XEndSegment ←NbStepsRights * Step
      YEndSegment ← NbStepsUp * Step
      DrawTill (XEndSegment, YEndSegment)
      InitW ← InitW + L
until InitW > SeqLength
```
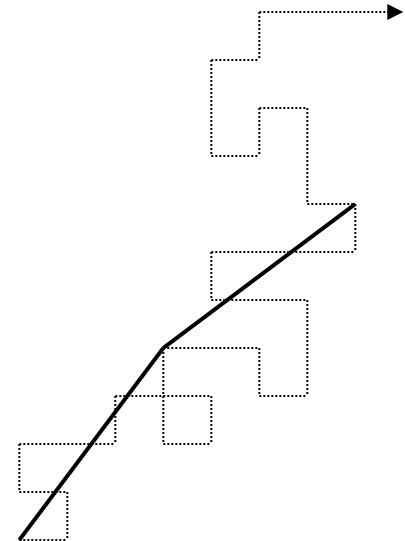
CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACAC

```
SeqLength, L, InitW, nbA,nbC,nbG,nbT, NbStepsRight, NbStepsUp: integer
XEndSegment, YEndSegment, Step: real
sequence: character string [1:*]
nbA,nbC,nbG,nbT ← 0
InitW← 1
repeat
      for i from InitW to min (InitW + L — 1, SeqLength) do
           case sequence [i] of
              "A": nbA ← nbA + 1
              "C": nbC ← nbC + 1
              "G": nbG ← nbG + 1
              "T": nbT ← nbT + 1
            endcase
      endfor
      NbStepsRight ← nbC — NbG
      NbStepsUp ← nbA — nbT
      XEndSegment ←NbStepsRights * Step
      YEndSegment ← NbStepsUp * Step
      DrawTill (XEndSegment, YEndSegment)
      InitW ← InitW + L
until InitW > SeqLength
```

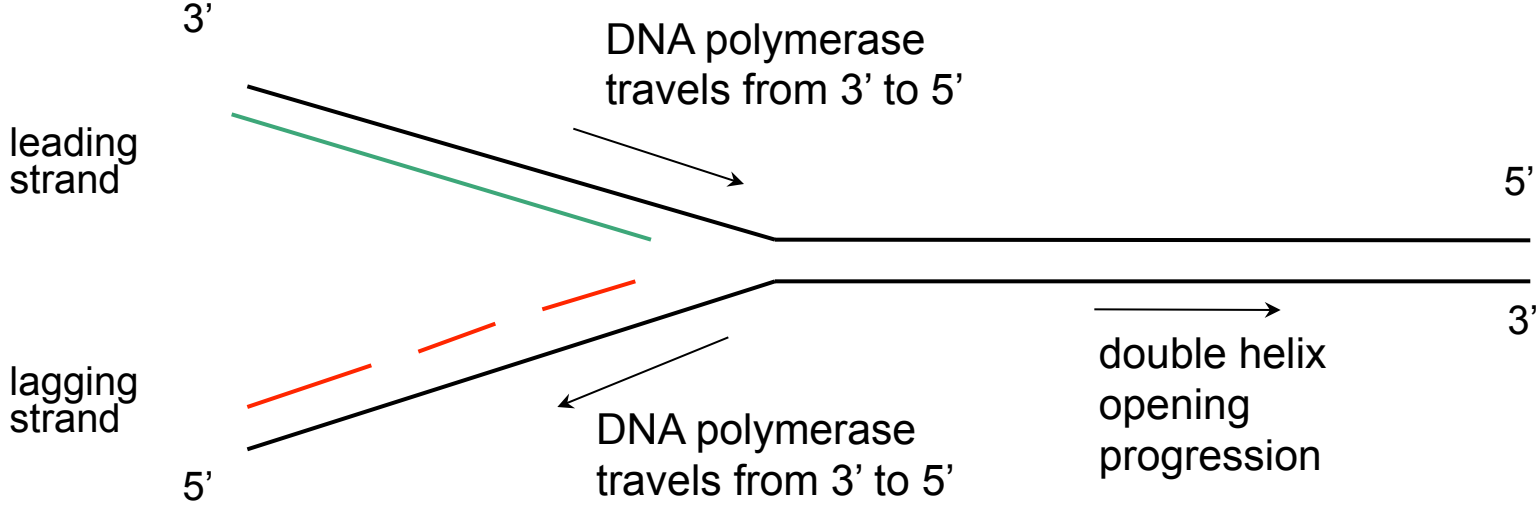**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACAC**
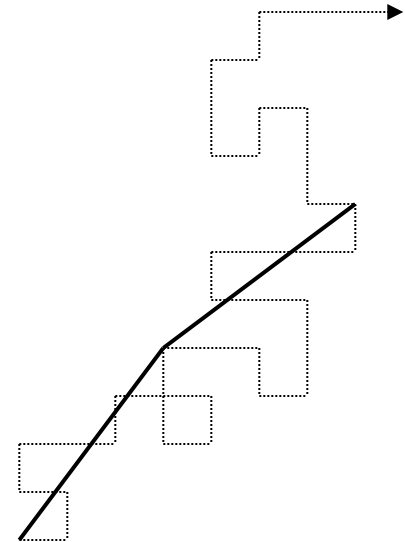
Running the algorithm on *Borrelia burgdorferi*

Beginning of the sequence; hence of the DNA walk

A
G
C
T

End of the sequence, and of the DNA walk
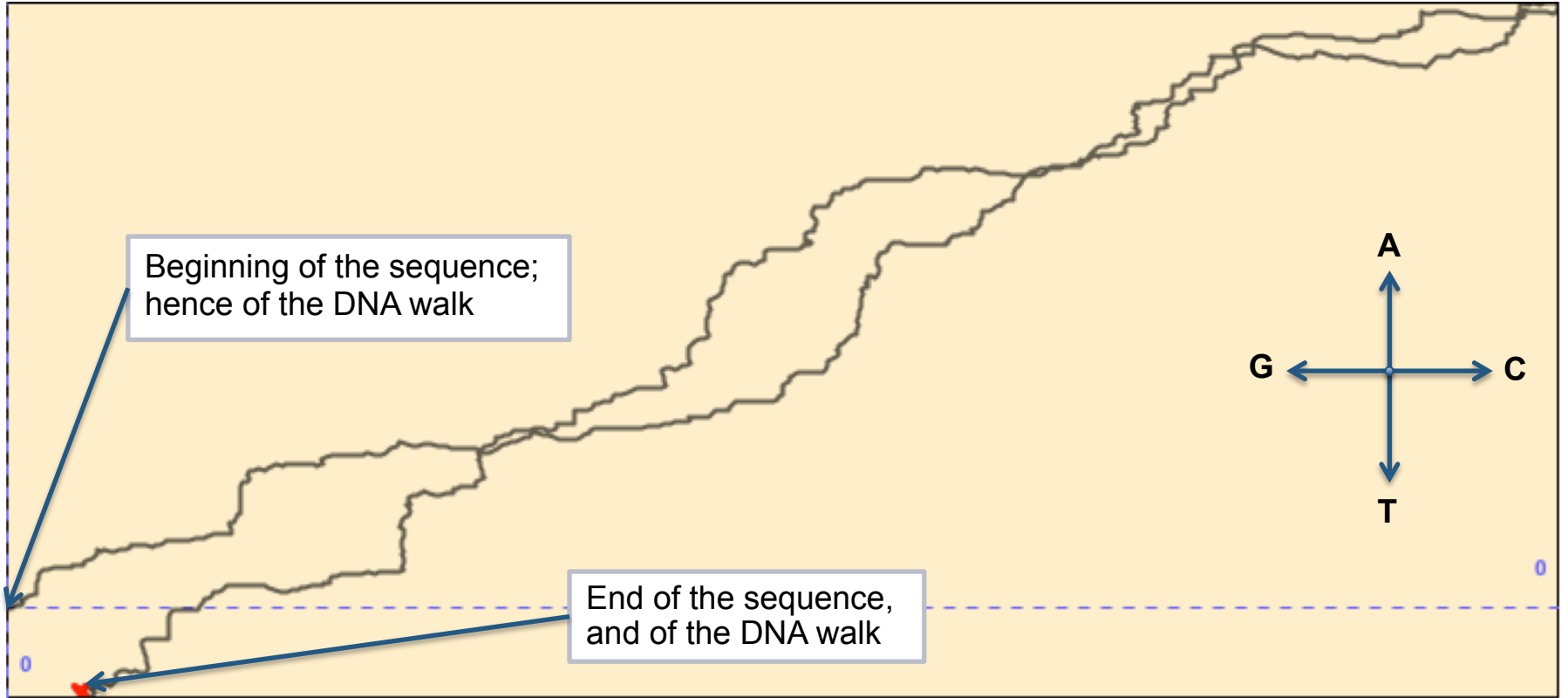
DNA walk and the origin of replication
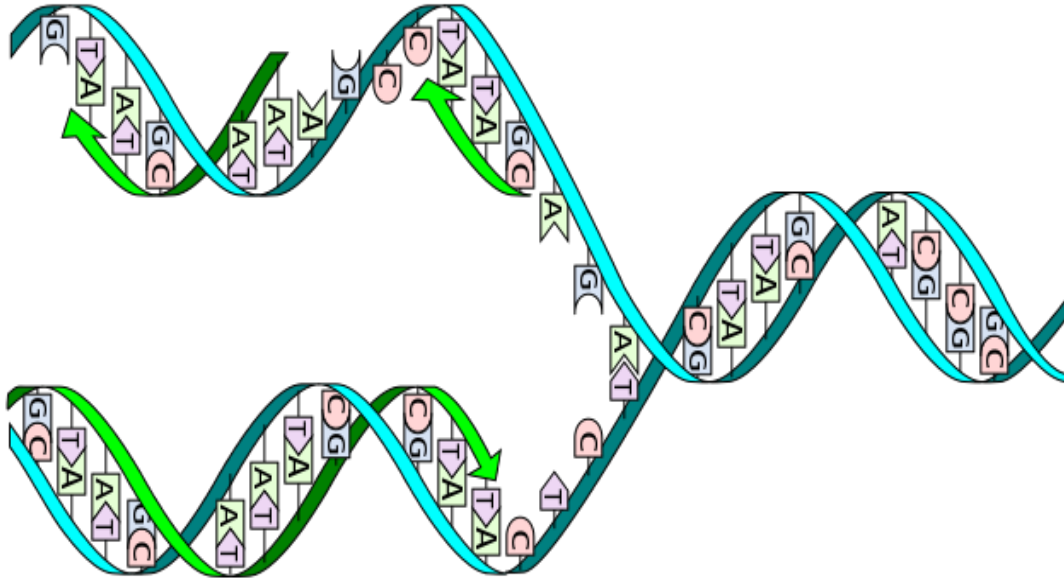
The DNA replication fork

leading strand

lagging strand

3'

5'

5'

3'

DNA polymerase travels from 3' to 5'

DNA polymerase travels from 3' to 5'

double helix opening progression

You cannot win every time you play!
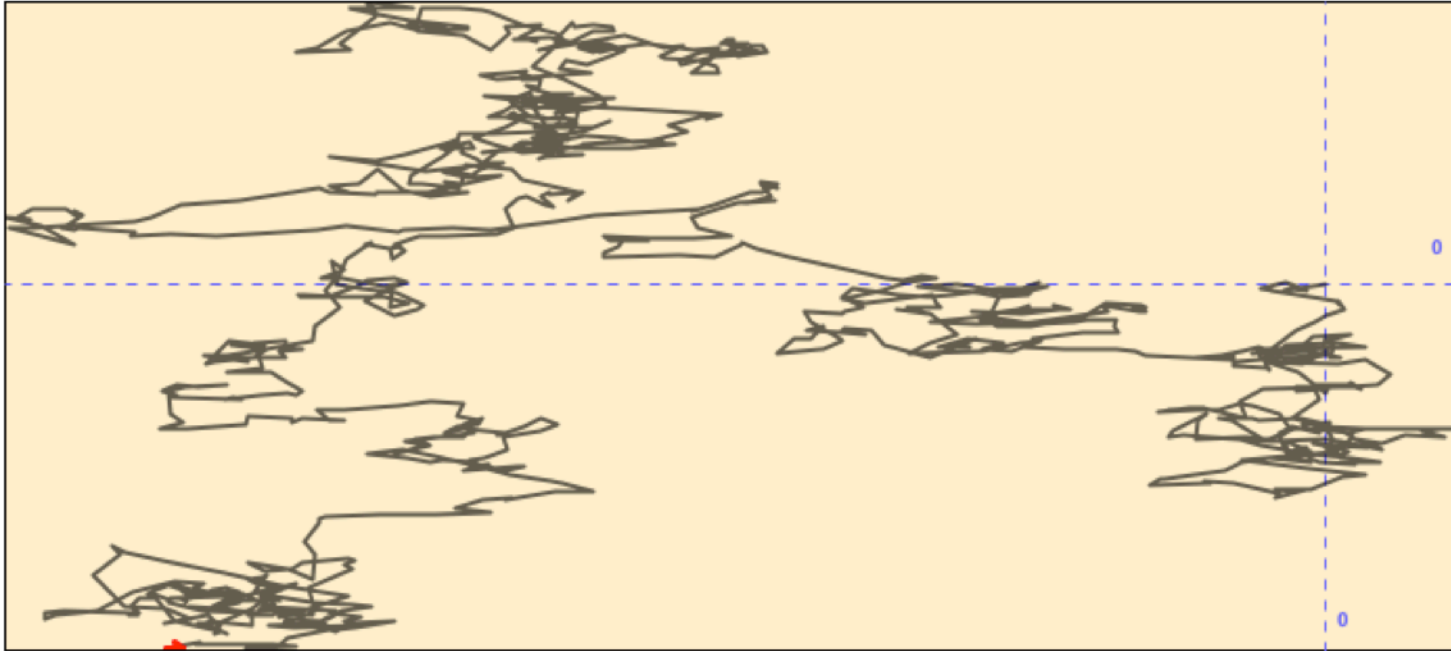
Synechocystis sp.

Pictures & movies : material licensing

p. 6, p. 7 : "DNA replication split" by I, Madprime. Licensed under CC BY-SA 3.0 via Wikimedia Commons