

# 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

# DNA walk

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAGAGTGTCTGATAGCAGC  
TTCTGAACTGGTTACCTGCCGTGAGTAAATTTAAATTTTATTGACTTAGGTCACTAAATACTTTAAACCA  
TATAGGCATAGCGCACAGACAGATAAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC  
ATTACCACCACCATCACCATTACCACAGGTAACGGTGC GGGCTGACGCGTACAGGAAACACAGAAAAAAG  
CCCGCACCTGACAGTGCGGGCTTTTTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTGAA  
GTTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTGCGGATATTCTGGAAAGCAATGCC  
AGGCAGGGGCAGGTGGCCACCGTCTCTGCCCCGCCAAAATCACCAACCACCTGGTGGCGATGATTG  
AAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCGGAACTTTT  
GACGGGACTCGCCGCCGCCAGCCGGGGTTCGCTGCGCAATTGAAAACTTTTCGTCGATCAGGAATTT  
GCCCCAAATAAAACATGTCTGCATGGCATTAGTTTGTGGGGCAGTGCCCGGATAGCATCAACGCTCGC  
TGATTTGCCGTGGCGAGAAAATGTGCATCGCCATTATGGCCGGCGTATTAGAAGCGCGCGGTCAACAACGT  
TACTGTTATCGATCCGGTCGAAAAACTGCTGGCAGTGGGGCATTACCTCGAATCTACCGTCGATATTGCT  
GAGTCCACCCGCCGTATTGCGGCAAGCCGCATTCCGGCTGATCACATGGTGCTGATGGCAGGTTTCACCG  
CCGGTAATGAAAAAGGCGAACTGGTGGTGCTTGGACGCAACGGTTCGACTACTCTGCTGCGGTGCTGGC  
TGCCTGTTTACGCGCCGATTGTTGCGAGATTTGGACGGACGTTGACGGGGTCTATACCTGCGACCCGCGT  
CAGGTGCCCGATGCGAGGTTGTTGAAGTGCATGTCTACCAGGAAGCGATGGAGCTTTCCTACTTCGGCG  
CTAAAGTTCTTACCCCCGCACCATTACCCCCATCGCCCAGTTCCAGATCCCTTGCCCTGATTAAAAATAC  
CGGAAATCCTCAAGCACCAGGTACGCTCATTGGTGCCAGCCGTGATGAAGACGAATTACCGGTCAAGGGC  
ATTTCCAATCTGAATAACATGGCAATGTTACGCGTTTCTGGTCCGGGGATGAAAGGGATGGTCGGCATGG  
CGGCGCGCGTCTTTGCAGCGATGTCACGCGCCCGTATTTCCGTGGTGTGATTACGCAATCATCTTCCGA  
ATACAGCATCAGTTTCTGCGTTCACAAAGCGACTGTGTGCGAGCTGAACGGGCAATGCAGGAAGAGTTC  
TACCTGGAAGTGAAGAAGGCTTACTGGAGCCGCTGGCAGTGACGGAACGGCTGGCCATTATCTCGGTGG  
TAGGTGATGGTATGCGCACCTTGCCTGGGATCTCGGCGAAATCTTTGCCGCACTGGCCCGCGCCAATAT  
CAACATTGTCGCCATTGCTCAGGGATCTTCTGAACGCTCAATCTCTGTGCGTGGTAAATAACGATGATGCC  
ACCACTGGCGTGC GCGTACTCATCAGATGCTGTTCAATACCGATCAGGTTATCGAAGTGTGTTGTGATTG  
GCGTCGGTGGCGTTGGCGGTGCGCTGCTGGAGCAACTGAAGCGTCAGCAAAGCTGGCTGAAGAATAAACA  
TATCGACTTACGTGTCTGCGGTGTTGCCAACTCGAAGGCTCTGCTCACCAATGTACATGGCCTTAATCTG  
GAAAAC TGGCAGGAAGAACTGGCGCAAGCCAAAGAGCCGTTAATCTCGGGCGCTTAATTCGCCTCGTGA  
AAGAATATCATCTGCTGAACCCGGTCATTGTTGACTGCACTTCCAGCCAGGCAGTGCGGGATCAATATGC  
CGACTTCTGCGGAAGGTTTCCACGTTGTCACGCCGAACAA

# Turn genetic sequences into music?

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCGCAGC



- Four letters, four frequencies only...

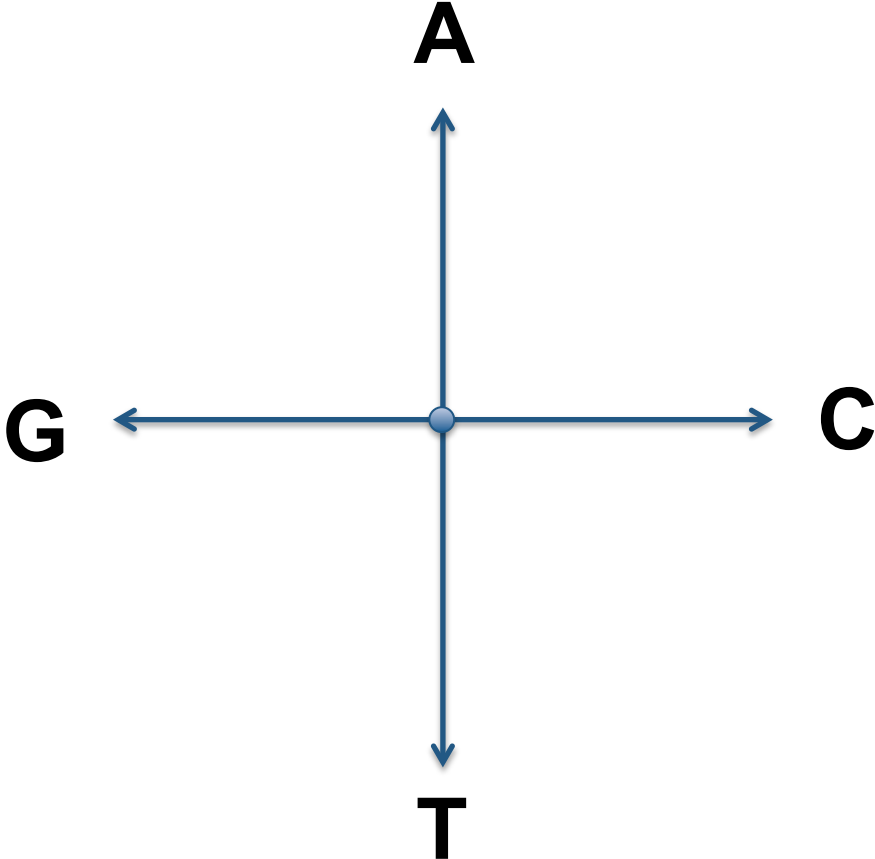
# Turn genetic sequences into music?

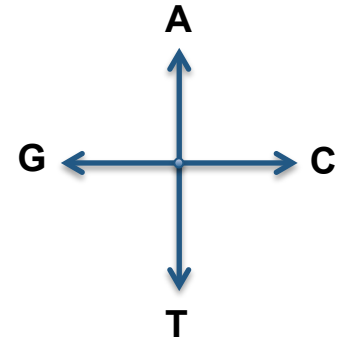
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCGCAGC



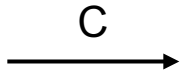
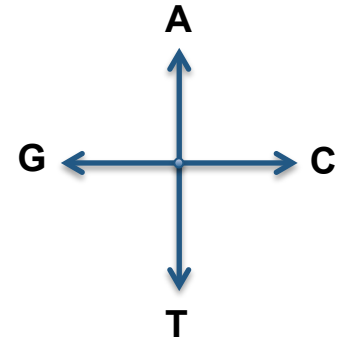
- Four letters, four frequencies only...
- Group the letters 3 by 3      AGC   TTT   TCA   TTC   TGA   CTG   CAA   CGG
  - $4 \times 4 \times 4 = 64$  different triplets

**Four letters? Four directions in the plane!**



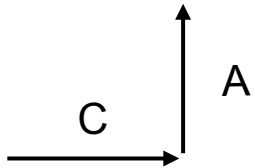
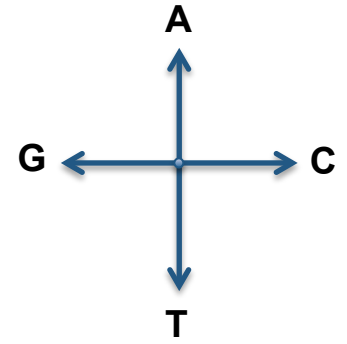


**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**

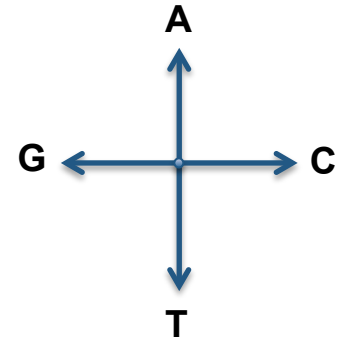
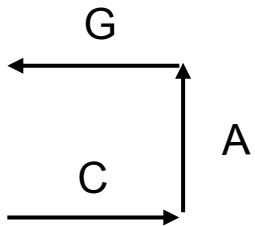


**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**

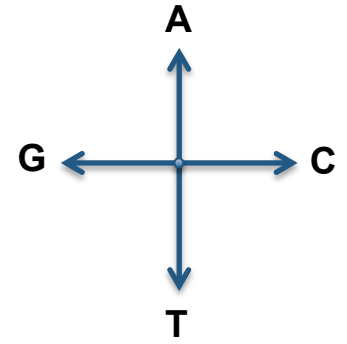
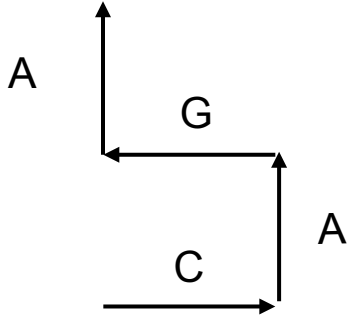




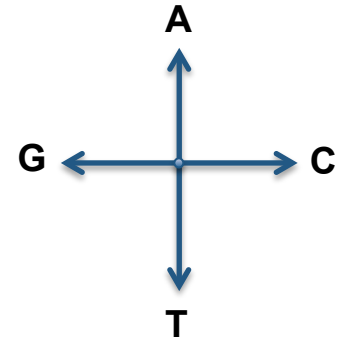
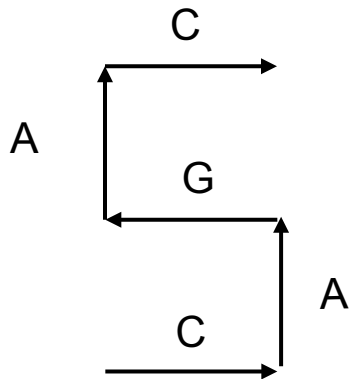
**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



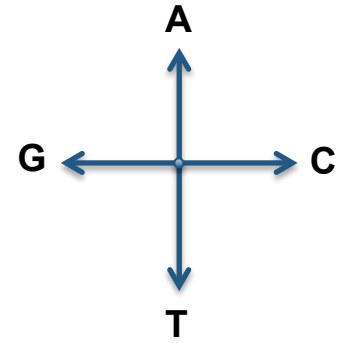
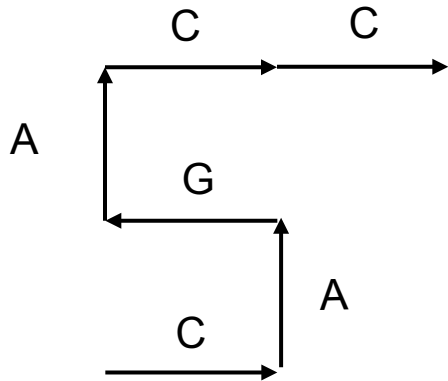
**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



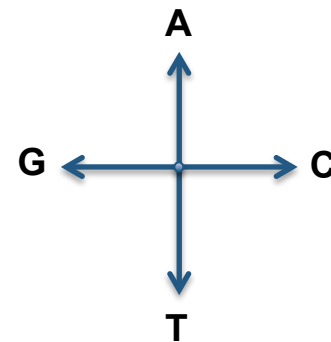
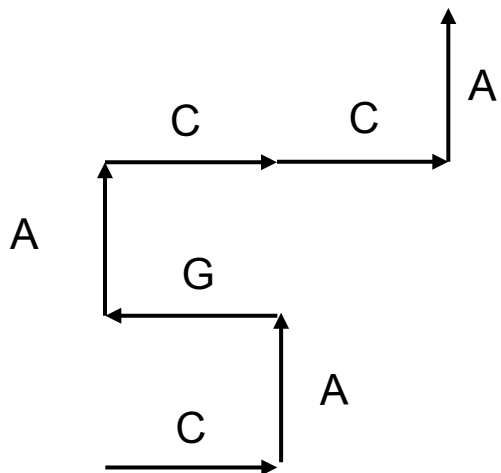
**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



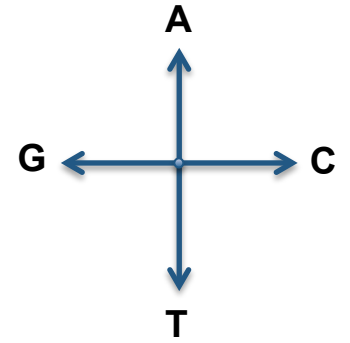
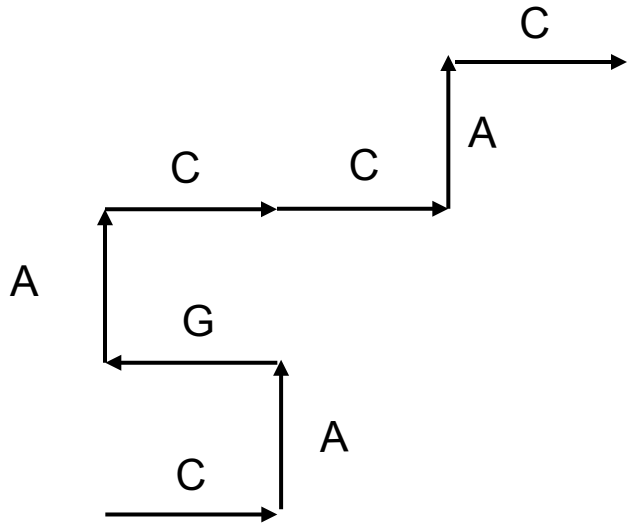
**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



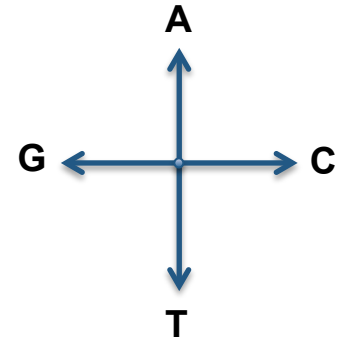
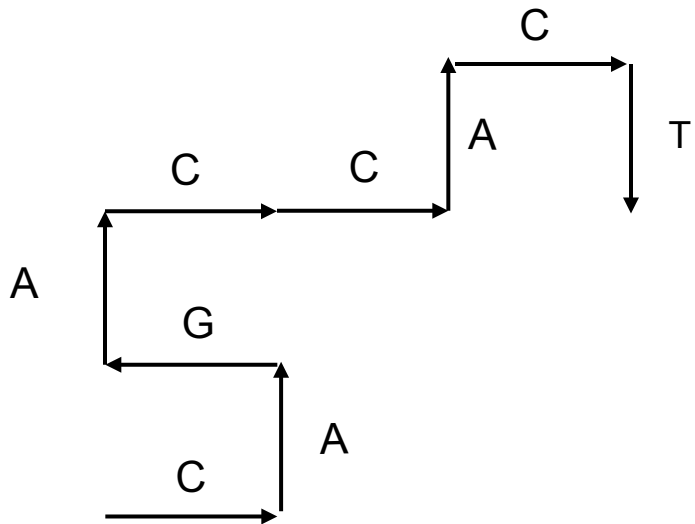
**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**

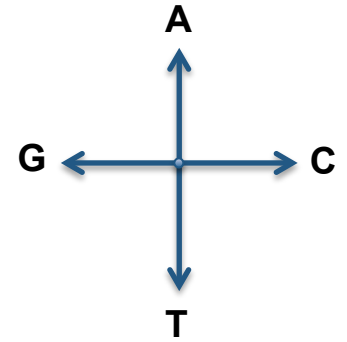


**CAGACCACTCAGACCTCAAGGACCCAGAAGTGAACACC...**



```
index: integer
sequence: character string [1:*]
index ← 1
repeat
  case sequence [index] of
    "A": drawUp
    "C": drawRight
    "G": drawLeft
    "T": drawDown
  endcase
  index ← index + 1
until sequence [index] = "*"

```



# 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?

# Pictures & movies : material licensing

p. 4, p. 5 : Cole SWE at en.wikipedia (Transferred from en.wikipedia) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], from Wikimedia Commons