BMEG3105 L4 by Prof. Yu LI

4.1. Sequencing importance

Non-invasive testing, e.g.: down syndrome

Used by 7 million of pregnant women every year

4.2. Recap

DP explores all choices and find optimal options

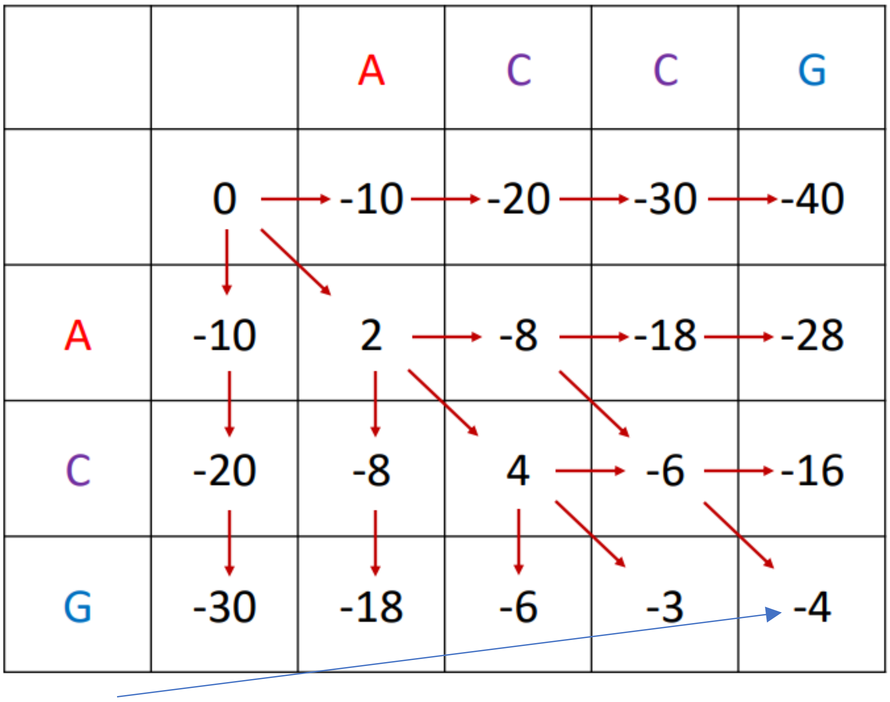
It divides and conquer the original problem for effectiveness and simplicity

4.3. DP table

Simplify reduction process

Obtain optimal alignment and optimal score

Example: DP with gap penalty = -10



		A	C	C	G
	0	-10	-20	-30	-40
A	-10	2	-8	-18	-28
C	-20	-8	4	-6	-16
G	-30	-18	-6	-3	-4

Optimal score = -4

Optimal alignment:

ACCG or ACCG
AC_G A_CG

4.4. Global and local alignment

Global: find optimal alignment among the whole sequence

Local: optimized matching subsequence

4.5. Reasons for mismatch and gaps

Mismatches: mutation, frame shifting
Gaps: indel, duplications

4.6. DP and Enumeration

Enumeration: calculates all possible results, time and resources consuming
DP: divide and conquer, filling the table
Big O notation with sequence length n in
  Enumeration: O(2^n)
  DP: O(n^2)

4.7. Software and tools

Webserver like EMBOSS: input the sequence and it will do the comparison, user friendly
Biopython: python library, allows more customized bioinformatic usage

4.8. Why sequencing

Better understanding of central dogma, phenotype and genotype
Discover hidden genetic information

4.9. Getting gene expression matrix

1. Read by sequencing
2. Map short read to genome
3. Count number of reads

4.10. Genome

Illumina sequencing length: ~200 bp
Human genome length: 3 x 10^9 bp