INTRODUCTION

How COMPUTER have become FUNDAMENTAL for MOL BIO

EXAM: 3 CFU 1h written

16/18 CROCETTE
3/4 OPEN QUESTIONS (5 LINES)

2 SEPARATED APPELLI can easily separate the exams

NO NEED OF PERFECT ENGLISH

---

really time consuming in the past, and fundamentally very slow

↓

Entered a phylosophycal approach to study in deep details genes

→ and all the things related to the gene

TECHNIQUES

- EXPRESSIVE CLONING
- PCR
- GEL ELECTROPHORESIS
- SOUTHERN / WESTERN BLOTs
- IMMUNOPRECIPITATION
- YEAST 2 HYBRID

studying a single gene in great details leads to a better understanding of the others genes

It's a LINEAR MACHINE → it will always respond in the same way

But there is all the rest that is involved with the genes that shall not be forgotten

GENOME WIDE / LARGE SCALE

≠ from SYSTEM BIOLOGY

↓

want to predict the behaviour of a system

→ very far away from doing it successfully

ex: gene · TP53

↓

encodes p53

So GENOME - WIDE studies

- How many genes are there, and where (GENOMICS)
- How many and which transcripts are produced (TRANSCRIPTOMICS)
- Which are the prot. transcripted? (PROTEOMICS)

↓

we can look at all of these at the same time

↓

mezzi di oggi

Molecular Biology and Genetics - TODAY

Genome-Wide Studies.

- How many genes are there, and where are they (Genomics)
- How many and which transcripts are produced (Transcriptomics)
- Which are the proteins encoded (Proteomics)
- What is the function of the genes/proteins (Functional Genomics) - classification of genes/proteins based on their function
- How the genome and its genes evolved over time, and its relationships with genomes of other species (Comparative Genomics)
- Characterize at the genome level interactions among genes and their products and their effect (Systems biology)

IF WE CAN SUM UP ALL THIS WE CAN UNDERSTAND THE BIGGER PICTURE

↓

AND KNOW WHERE TO LOOK

Modern molecular biology and genetics

Genome and epigenome

- DNA sequencing and assembly of the complete genome
- RNA sequencing and annotation of the complete transcriptome
- Protein sequencing and structure
- Prediction of gene/transcript/protein function and interactions
- Simultaneous genome-wide measurement of genes' transcript levels
- Genome-wide epigenomic maps (DNA methylation, nucleosome positioning, histone modifications, DNA accessibility, etc.)

-ome
& that comprend all

Modern molecular biology and genetics

Genome and epigenome

- Whole-genome gene annotation starting from DNA and RNA sequences
- Genome-wide protein annotations, from mRNA sequences
- Genome-wide functional annotation of genes and protein
- Identification of genes with similar expression patterns, which can be tissue- condition- disease-specific
- Identification of genetic and epigenetic factors regulating the expression of genes with similar expression patterns

What is BIOINFORMATICS?

Scope of bioinformatics

- Storage and retrieval of biological data
- Sequence analysis
  - Sequence alignments, database searches, motif detection
- Genomics
  - genome annotation,
  - comparative genomics
  - genetics
- Phylogeny
- Functional genomics
  - Transcriptome, proteome, interactome
- Analysis of biochemical networks
  - metabolic networks, regulatory networks
- Systems biology
  - Modelling and simulation of dynamic systems
- Molecular structures
  - visualization and analysis, classification, prediction
- ...

Biology is a comparative science

Genomes (at NCBI)

→ ≠ SPECIES

- 8543 complete EUKARYOTIC genomes
- 204452 PROKARYOTIC Genomes
- 29960 VIRAL genomes
- 16474 PLASMIDS
- 12253 ORGANELLE genomes

- These are ALL underestimates as much data is never submitted to public databases for various reasons

SEQUENCING

Biological sequences can be read, and the decreasing cost of it has permitted the completion of hundreds of whole genome sequencing projects

its easier faster and cheaper to sequence DNA instead of PROT.

We can't directly sequence RNA

↓

we can reverse it in DNA and sequence the DNA

DNA SEQUENCING METHODS

- SANGER
  - accurate
  - up to 1000 nt
  - single clones or PCR products

NO MIXED!

"NEXT SEQUENCING METHODS"
- new technologies (2006-2016)
  - less accurate
  - shorter reads (< 300 nt)
  - cheap
  - high throughput seq of MIXED DNA

Sequencing

- Biological sequences (nucleotides or amino acids) can be read
- The (ever decreasing) cost of sequencing has permitted the completion of (now) hundreds of whole-genome sequencing projects, from viruses and bacteria to higher eukaryotes and human
- Now we are in the "next-generation" sequencing era, based on high-throughput-sequencing" techniques and platforms

DNA sequencing methods

- Sanger sequencing: → takes advantage of the basic DNA REPLICATION
  - Invented by Fred Sanger in 1970s
  - Accurate
  - Reads of up to 1000 nucleotides
  - Single clones or PCR products sequenced

- "Next Generation Sequencing"
  - New technologies (2006-2016)
  - Less accurate
  - Shorter reads (<300 nucleotides)
  - Very cheap, high throughput sequencing of mixed DNA fragments

For Sanger sequencing we need:

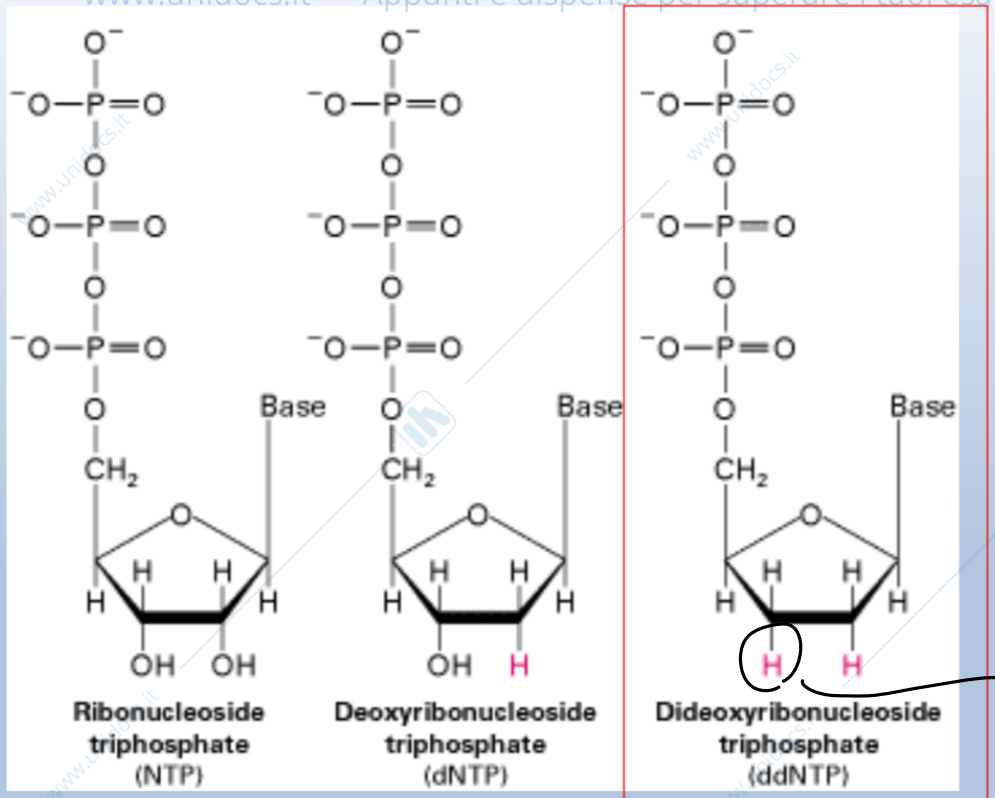
1)DNA template

2)A primer for DNA synthesis (as in PCR, complementary to template)

3)DNA polymerase

4)Deoxynucleoside triphosphates and deoxyNTPs

Dideoxynucleotide triphosphates



RNA DNA

NTP, dNTP and ddNTP (chain terminators)

component of DNA

dNTPs

ddNTPs

when inserted in a sequence occurs termination because there is not a 3' OH to which the next nucleotide can bond

↓

STOP

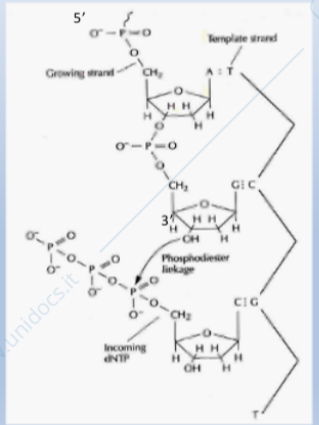
Sanger dideoxy sequencing–basic method

Single stranded DNA

a) Anneal a primer (complementary to a known part of the sequence (we often sequence DNA that has been CLONED into a plasmid vector – we know the sequence of the cloning site near where the "insert" starts. Otherwise if we have a PCR product, we know the sequence of the PCR primers))

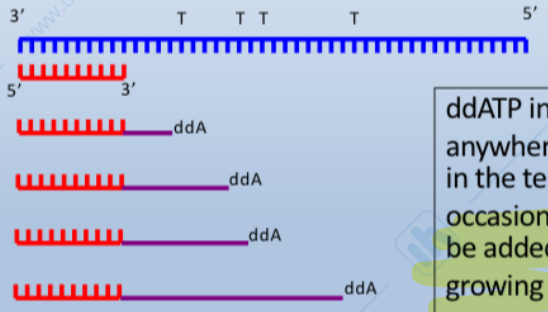
Sanger dideoxy sequencing: basic method

b) Extend the primer with DNA polymerase in the presence of all four dNTPs, with a limited amount of ONE dideoxy NTP (ddNTP)



Direction of DNA polymerase travel

Sanger dideoxy sequencing: basic method

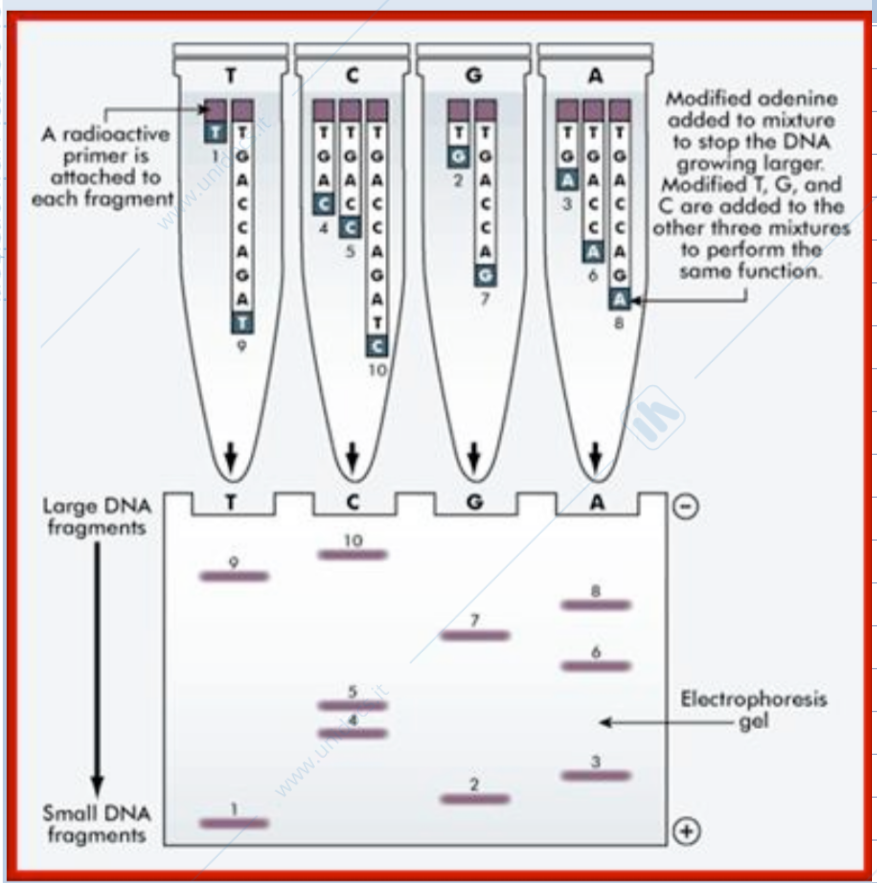


ddATP in the reaction: anywhere there's a T in the template strand, occasionally a ddA will be added to the growing strand

bc ddNTP is present in a limited amount, and only in one form

How to visualize DNA fragments?

- **Radioactivity**
  - Radiolabeled primers (kinase with ^{32}P)
  - Radiolabelled dNTPs (gamma ^{35}S or ^{32}P)



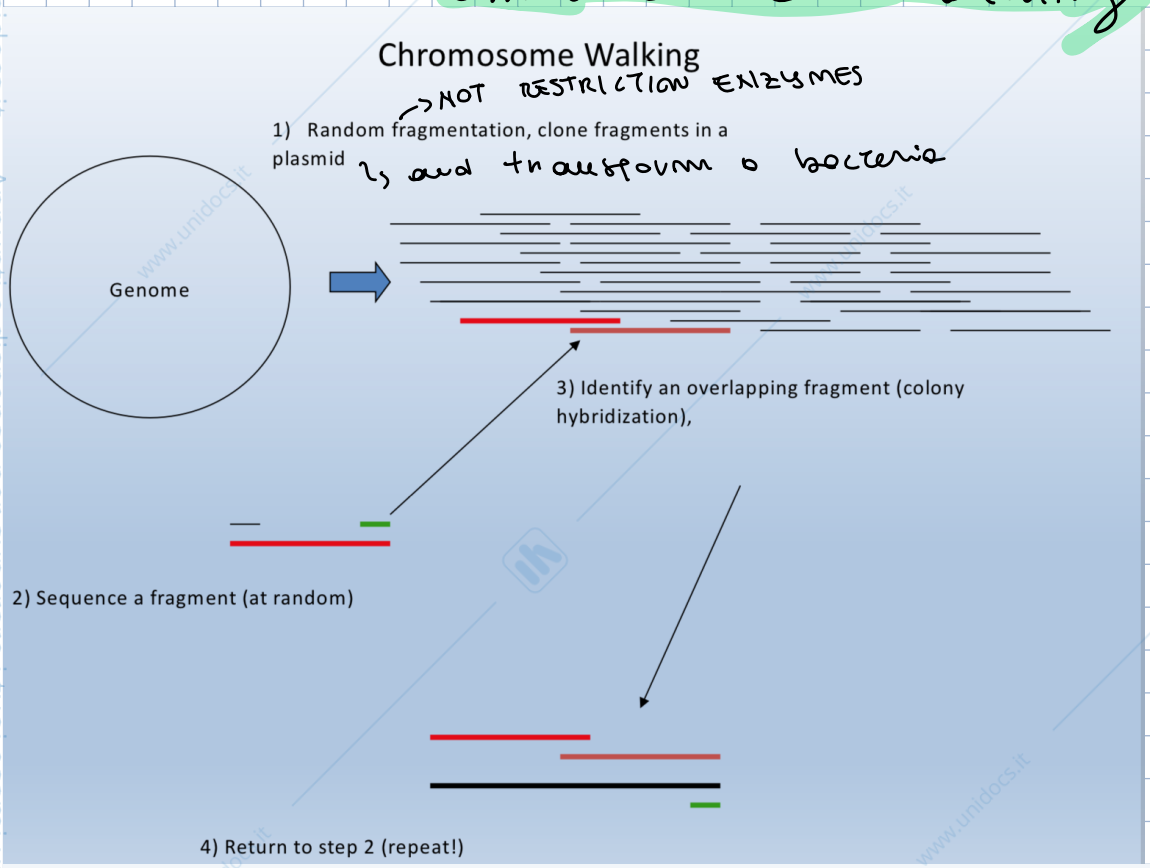
because it was tedious to do it

at first it was sequenced only small genomes

-> used to use

↓

Chromosome Walking → very slow



Divide and Conquer

For larger (but still small) genomes like EBV, or the chloroplast genome of Arabidopsis, it would take a long time to "walk" along the whole genome.

The chosen strategy was to divide the genome into RESTRICTION fragments and "walk" each one in parallel

EBV:

c. 40 BamHI restriction sites

each fragment cloned into a plasmid vector and sequenced by walking independently

Repeat with a different restriction enzyme to place the fragments in the correct order