Bioinformatics today and tomorrow

The last lecture

Today

REAL-LIFE PROBLEMS

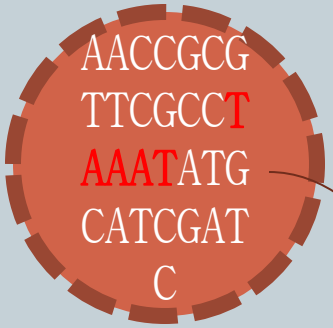
Living cells as computational devices

Code repository

```
public static void main (LIFE)

X=input()

If X==A
   protein A=new Protein (TAAATA…)
```

Program execution

Living cells as computational devices



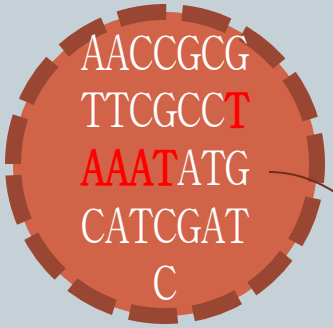
Code repository

```
public static void main (LIFE)

X=input()

If X==A
   protein A=new Protein (TAAATA…)
```

Program execution

Working copy of the code

Living cells as computational devices



Code repository

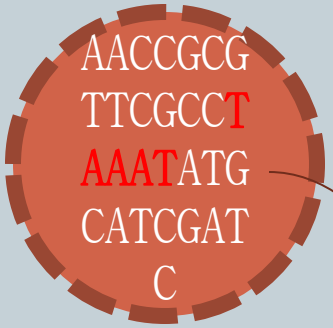
```
public static void main (LIFE)

X=input()

If X==A
   protein A=new Protein (TAAATA…)
```

Program execution

Working copy of the code

Output protein sequence

Living cells as computational devices



Code repository

```
public static void main (LIFE)

X=input()

If X==A
   protein A=new Protein (TAAATA…)
```

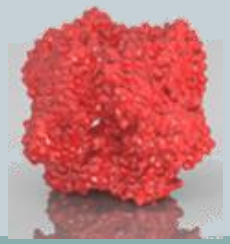
Program execution

Working copy of the code

Sequence-dependent folding

Output protein sequence

A H W A H P P M T U V A T M



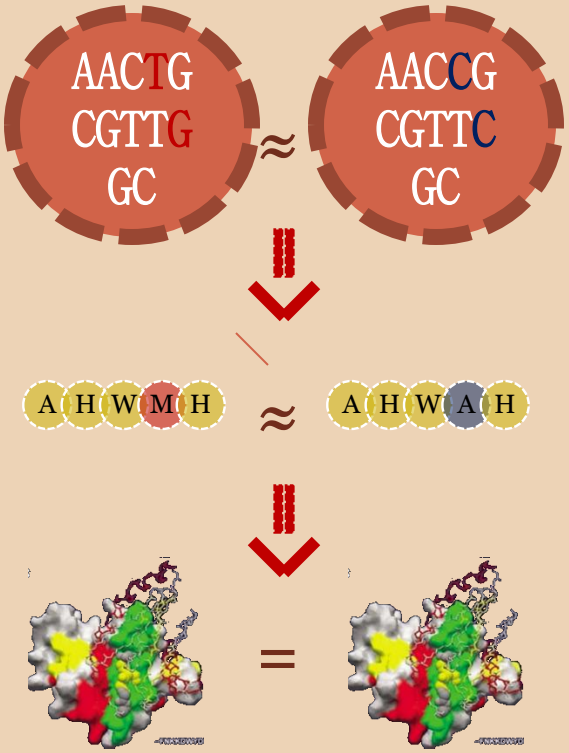
Digitalization of the molecular code

A *new type of digital data*

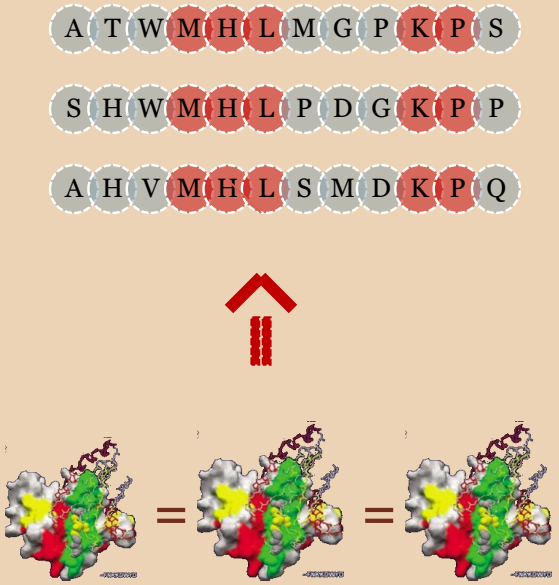
- Contiguous strings
- Misspellings
- Massive

Logical basis for biological sequence analysis

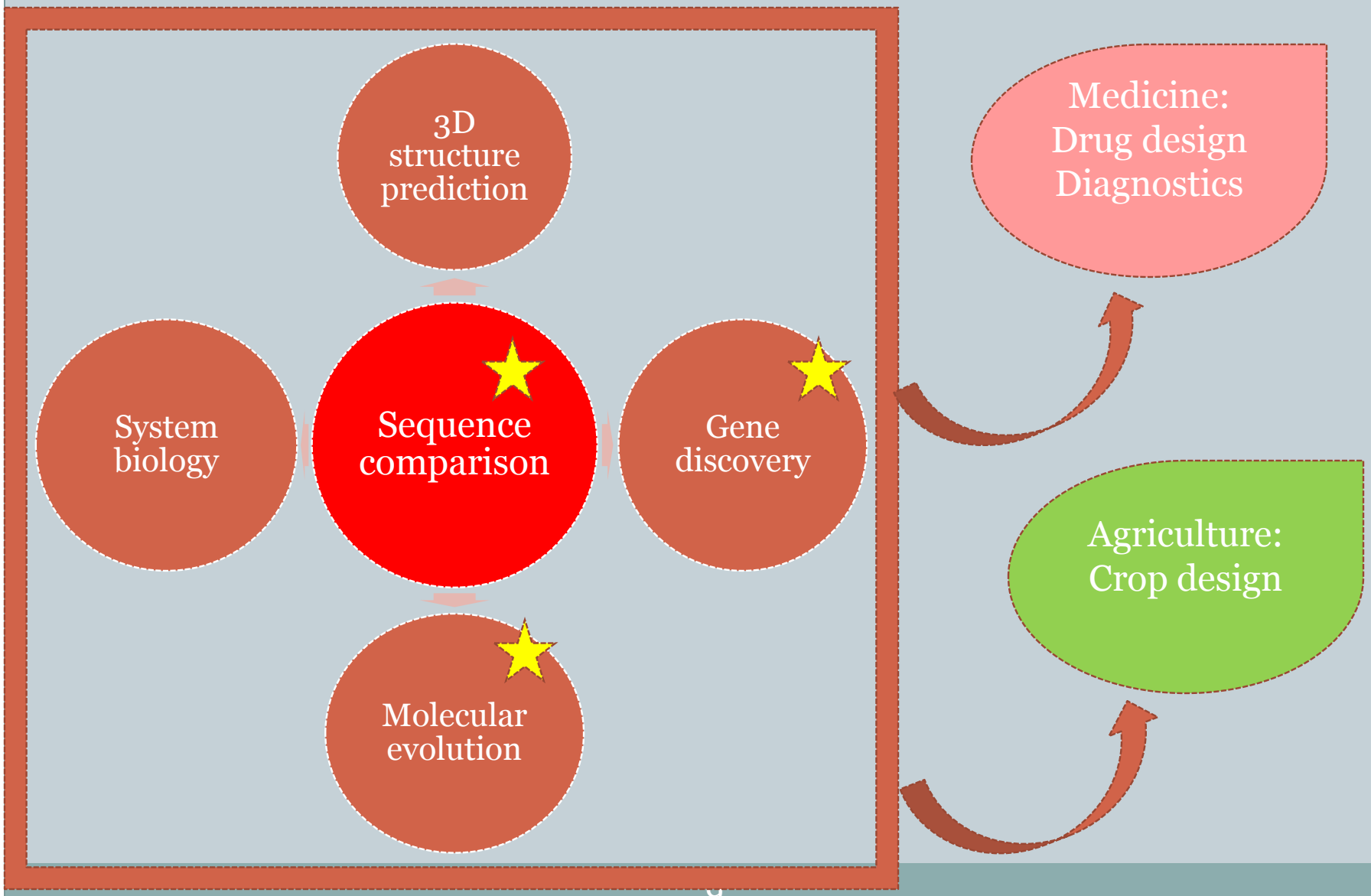
Similar sequences – similar function



Similar function – partly similar sequences

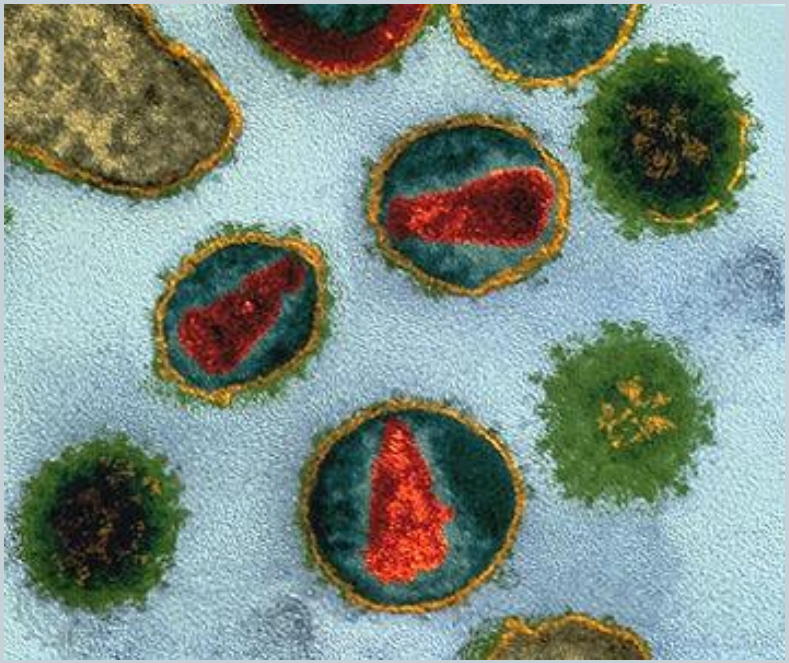


Bioinformatics at a glance



Problem 1.HIV virus: high mutation rate

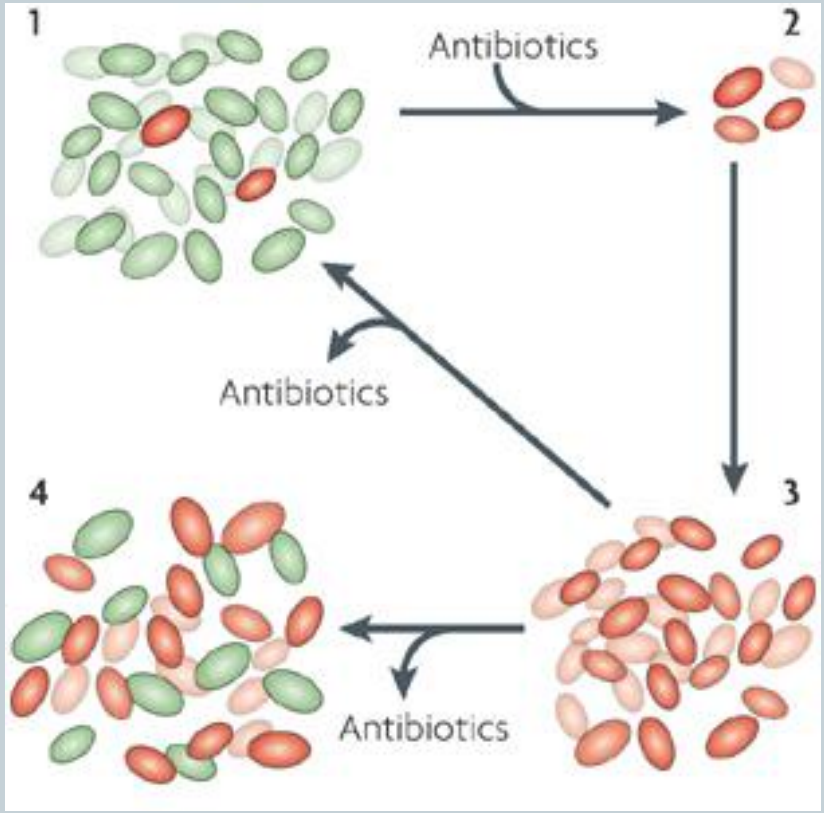
Subject: The plan - Re: request for alignment of HIV sequences



```
TNELVLDLXLCLLLHKMLSLXLYVYFFLYXCWXLXG
TMCLS-XLLLFFLLHEMLSLGL-IYFLXYW-GGXCN-
TMCLSYISWLFVLSYK-LSLGPIGTLFLVVLWGPGT
TMCLSYISWLLVLSYKLLSLGPIGCLFLVVLWGPGT
TMCLSYISWLFVLSYK-LSLGPICTLFLSVLWFPGT
TMCLSYISCYLSYCMKCSPW-XYMYALXYIFVGXC
TVCFSYISYYFSTCIKCPPW-XYMYASYCIGXGSCT
TMCLS ISCXLSCHINCSPW-SYRHXFPCXXIGXCT
TMCLSYISCXLSCHINCSPW-SYRDXFPCAXIGXCT
-TIWSYISFYSSCCMKCSPWXLYVLSFFXYWCWVX
TVCLSYISYYFSWLYTILSL-VLELPFFLXWXWVLY
TVCLSYISYYFSWLLTILSL-VLELDFFLXWX-VLYN
TMCLSYSXYGLLVHTILVP-XLYVHLFLYCCWVLYX-
-----IFPFIPPVA-XALPXPICIILLCIWLLGLVQLISTX
```

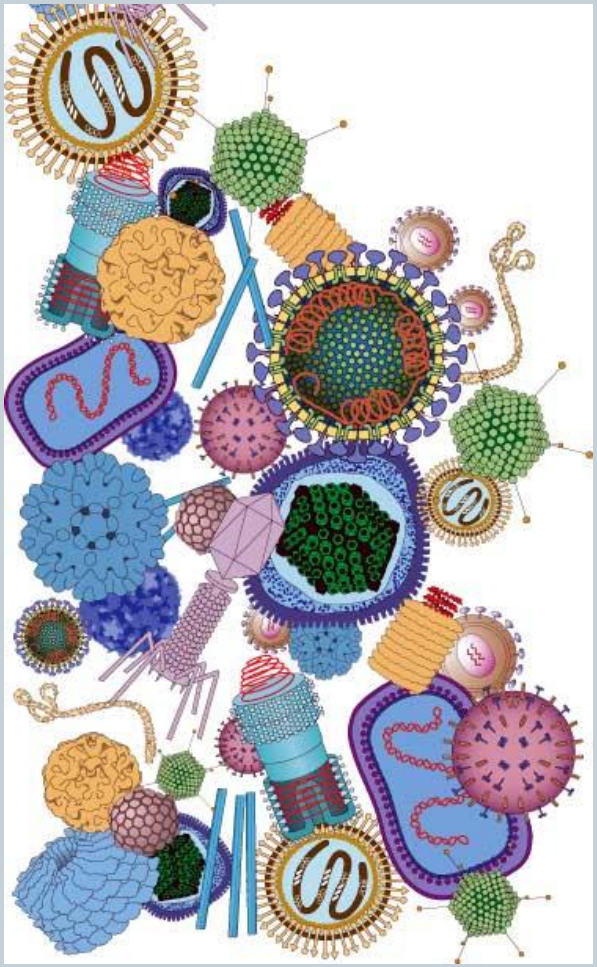
Find local regions of high similarity to design HIV drugs

Problem 2 (similar). Antibiotics resistance



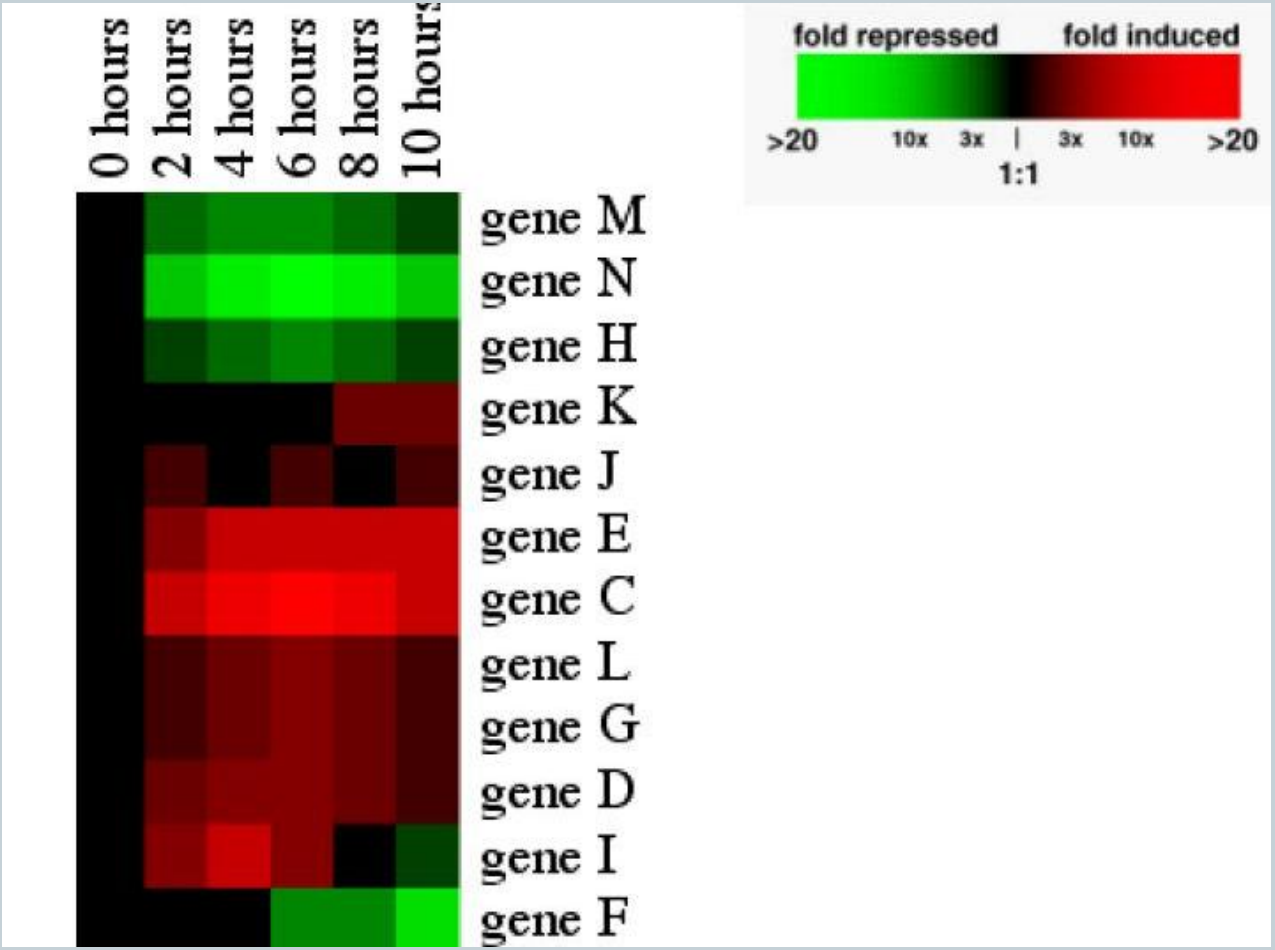
How to design a new type of antibiotics which will kill all the pathogenic bacteria, and no mutant strain will survive and proliferate?

Problem 3. Sequence redundancy in viral databases



How to remove the redundant sequence information from the database of viral genomes?

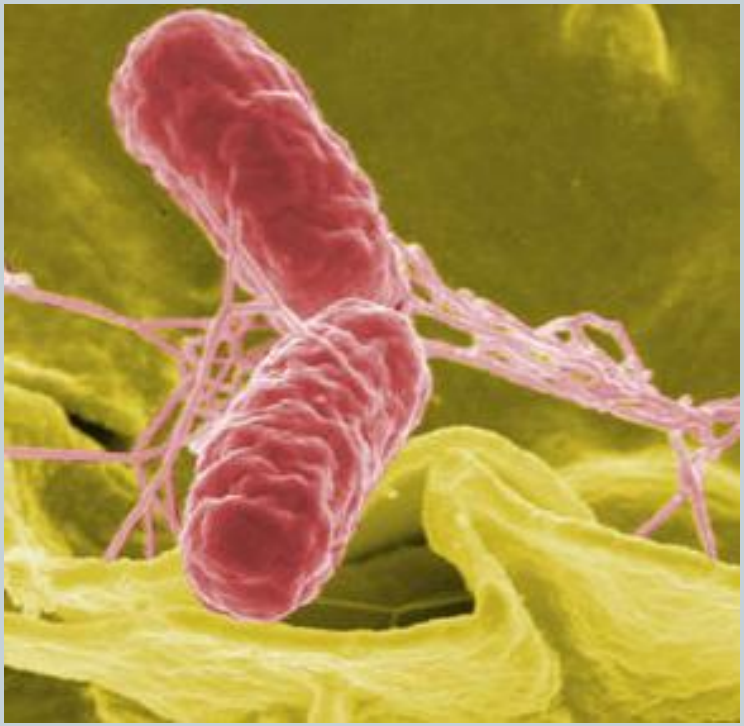
Problem 4. Expression patterns of genes



How to identify the group of genes with similar expression patterns?

Time series comparison

Problem 5. Lethal E.coli



O157-H7 modification of E. Coli (in undercooked beef) causes occasionally lethal hemorragic colitis

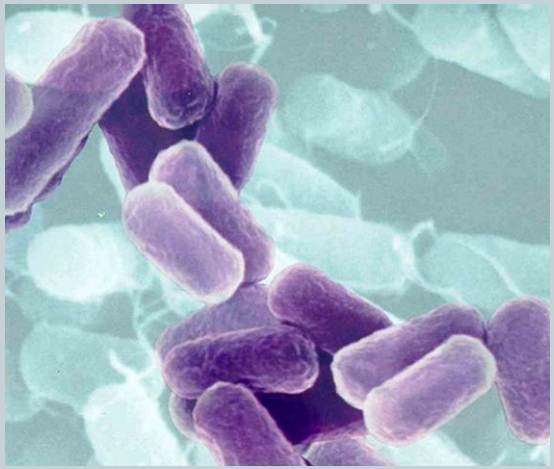
The pathogenic strain contains O-islands, and the normal strain contains K-islands, with different frequency of nucleotides.

How to detect an early mutation of a normal E. coli into a pathogenic strain?

Problem 6. A 3-domain system

Prokaryotes	Bacteria	Eubacteria	Bacteria
		Archaebacteria	Achaea
Eukaryotes	Protista	Protista	Eukarya
	Fungi	Fungi	
	Plantae	Plantae	
	Animalia	Animalia	

How to group them?



Bacteria

Achaea

Eukarya

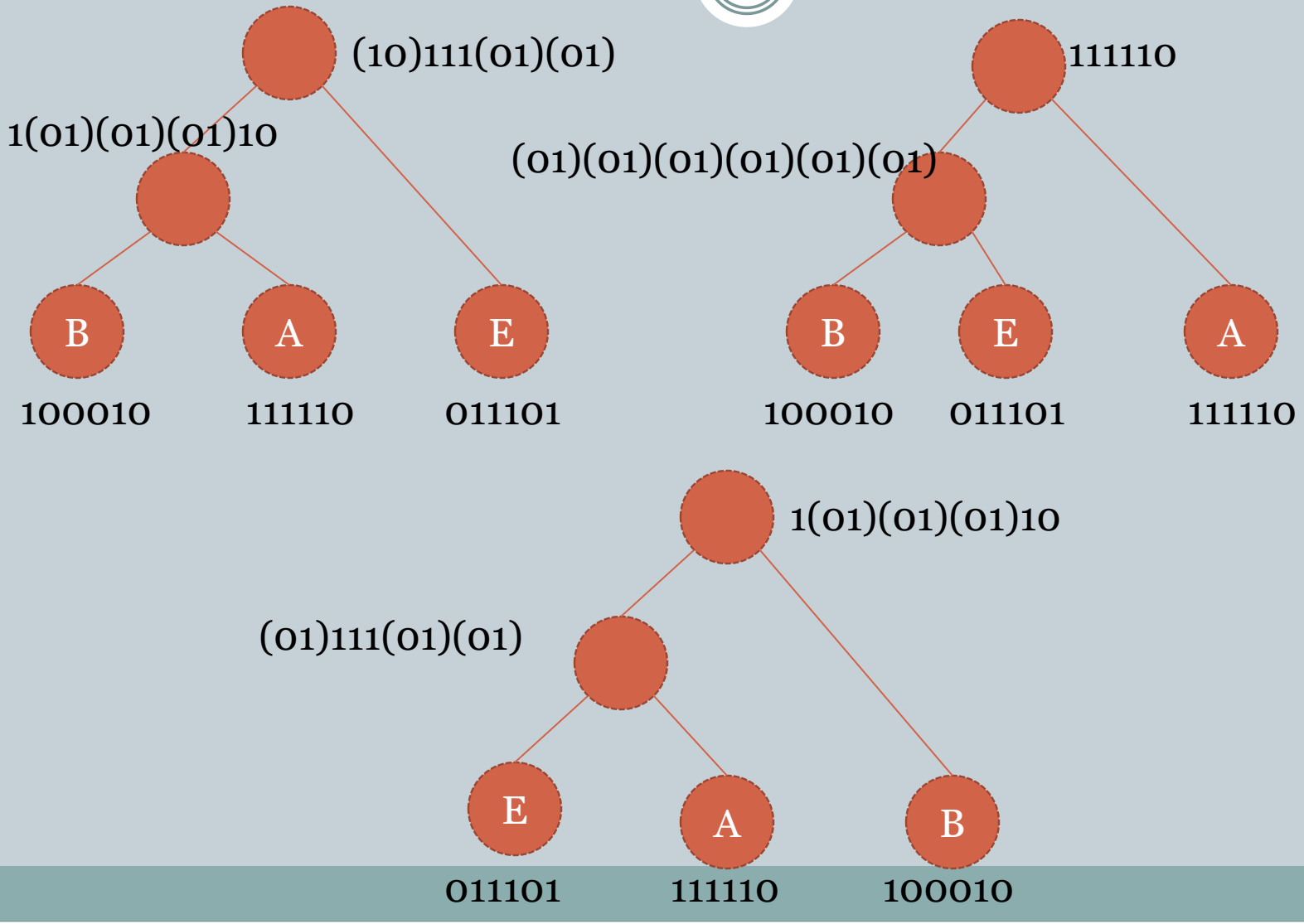
Table of binary molecular attributes

	ribosome	RNA-polymerase	exons	First t-RNA	Operon	Nucleus in membrane
B	70S	1	rare	Formil-methyonine	yes	no
A	70S	several	yes	methionine	yes	no
E	80S	many	yes	methyonine	no	yes

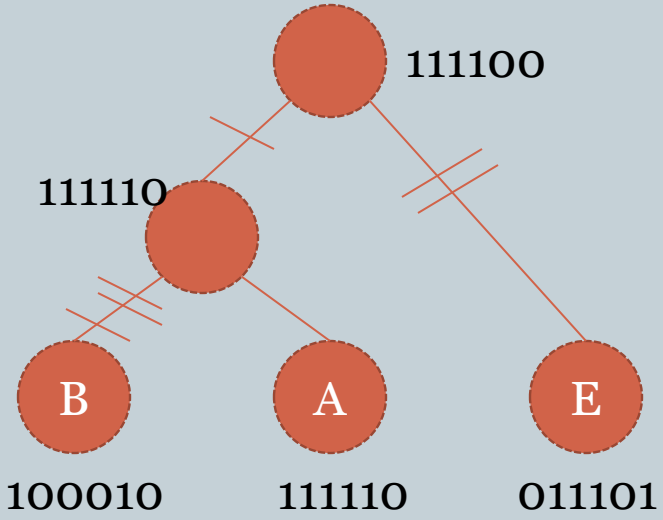
Table of binary molecular attributes

	C1	C2	C3	C4	C5	C6
B	1	0	0	0	1	0
A	1	1	1	1	1	0
E	0	1	1	1	0	1

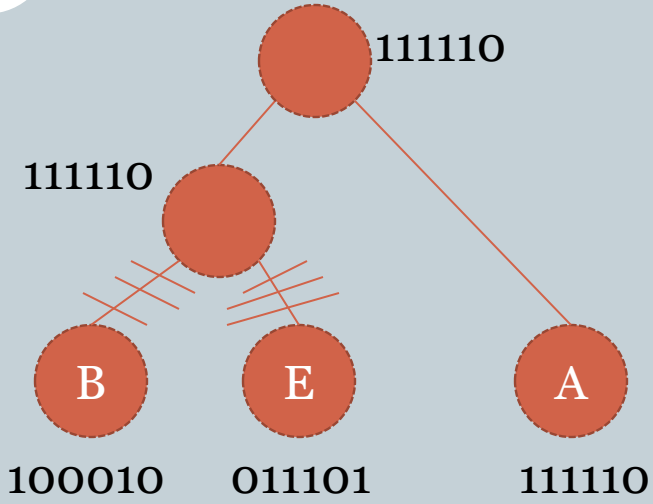
The most parsimonious tree?



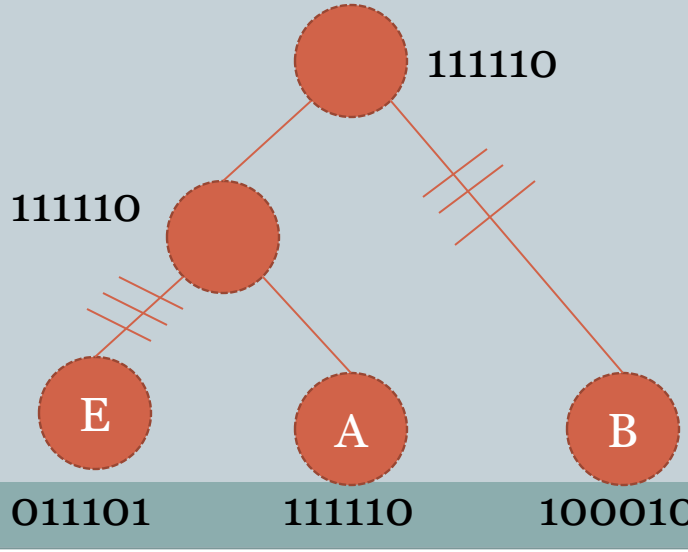
The most parsimonious tree?



L(T)=6



L(T)=6



L(T)=6

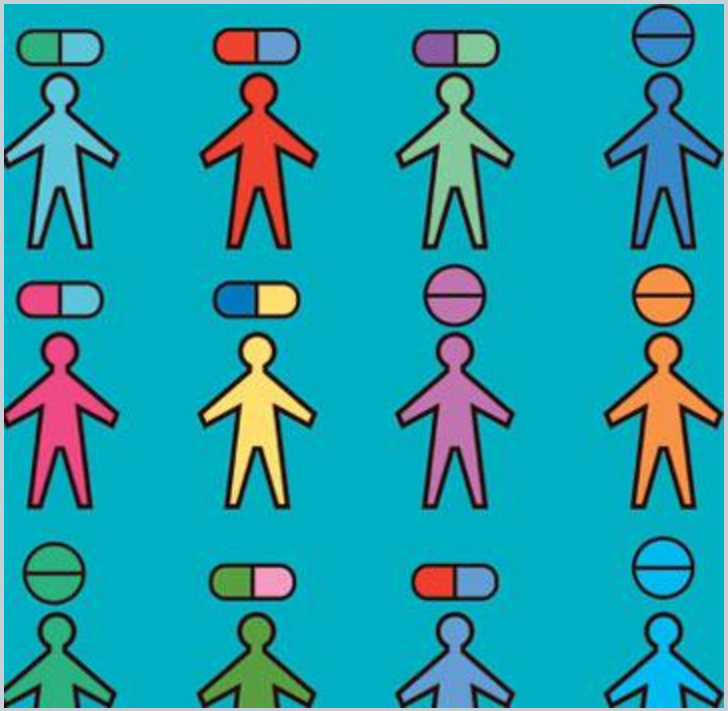
Tomorrow

SCIENCE FICTION?

Everybody gets his genome sequenced

today	tomorrow
• 10,000 $	• 500$

Personalized medicine



THE GENETICALLY DETERMINED REACTION TO DRUGS WILL BE DEDUCED FROM THE COMPARISON OF YOUR GENOMIC SEQUENCE WITH THE SEQUENCE DATABASES AND YOU GET THE TREATMENT WHICH IS THE BEST FOR YOU

Global test for all genetic diseases

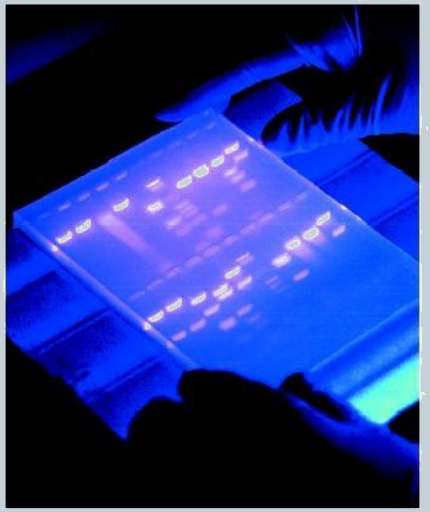
ARE YOU A CARRIER OF A GENETIC DISORDER?

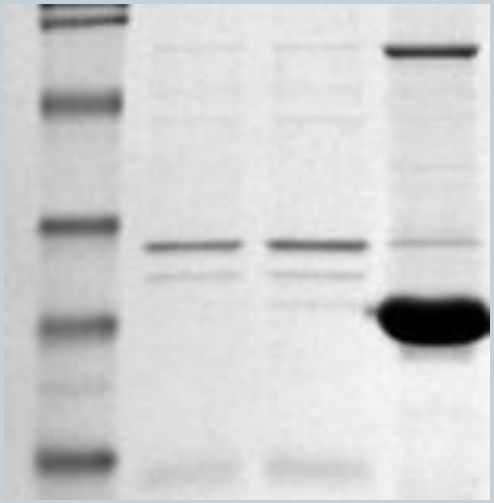
SELECTION OF A HEALTHY GENOTYPE AMONG ALL FERTILIZED EGGS

DNA hybridization

25

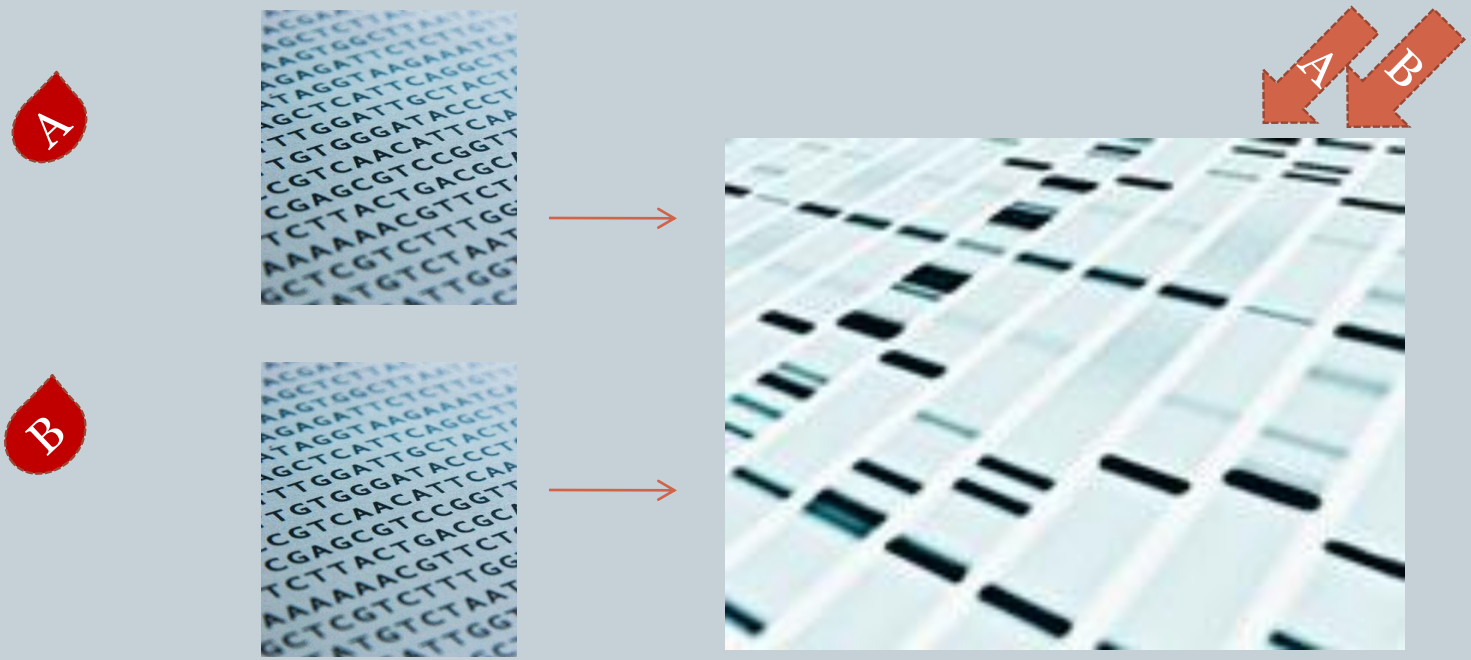
TODAY





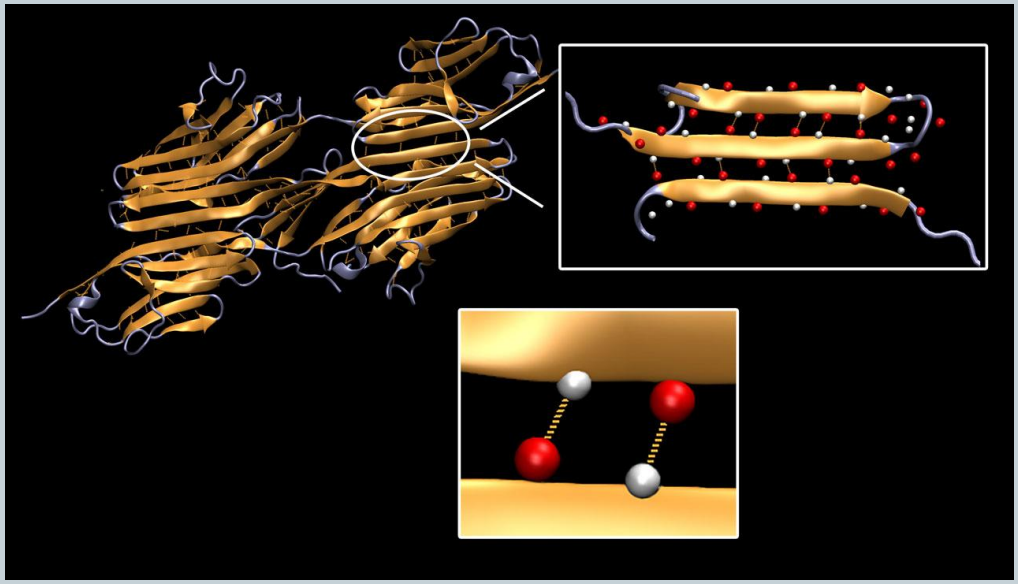
In silico DNA hybridization

TOMORROW



Imagine: engineering projects at UVic

public static void main (SPIDER_SILK)



IMPLEMENT A LIVING SYSTEM WHICH PRODUCES THE BIO-DEGRADABLE PLASTIC

IMPLEMENT BACTERIA WHICH PRODUCES THE SPIDER-SILK PROTEIN

…