BIOINFORMÁTICA

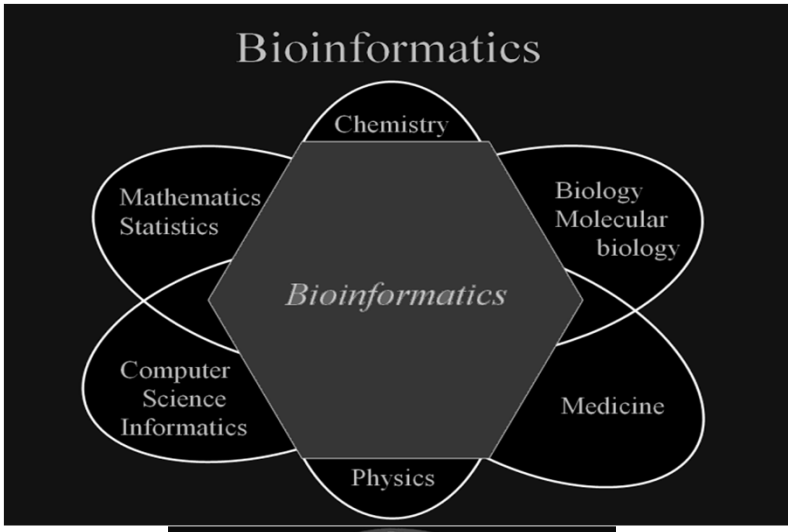
Vicente Arnau Llombart

http://www.uv.es/~varnau/bioinformatica.pdf

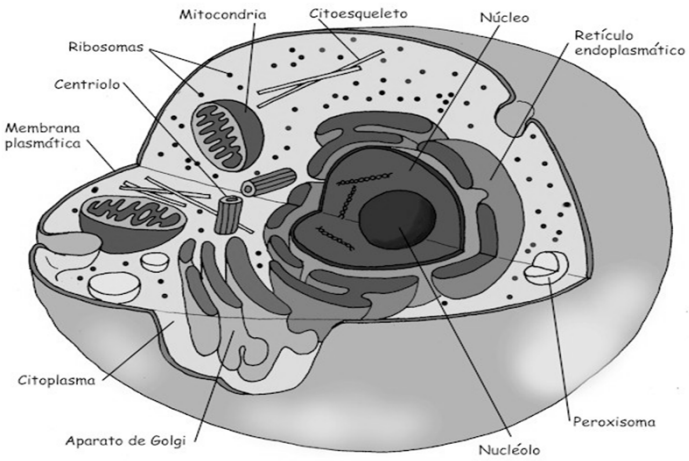
E-mail: Vicente.Arnau@uv.es

¿Que es la BIOINFORMÁTICA?

BIOLOGIA MOLECULAR ↔ INFORMÁTICA



La Celula

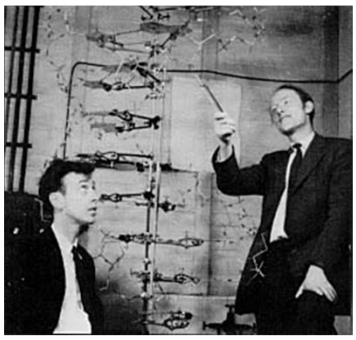


C. EUCARIOTA Humana posee ➔ 3 Gbases en el NUCLEO y 16 Kbases en las MITOCONDRIAS

VNIVERSITAT DE VALENCIA

El Genoma Humano

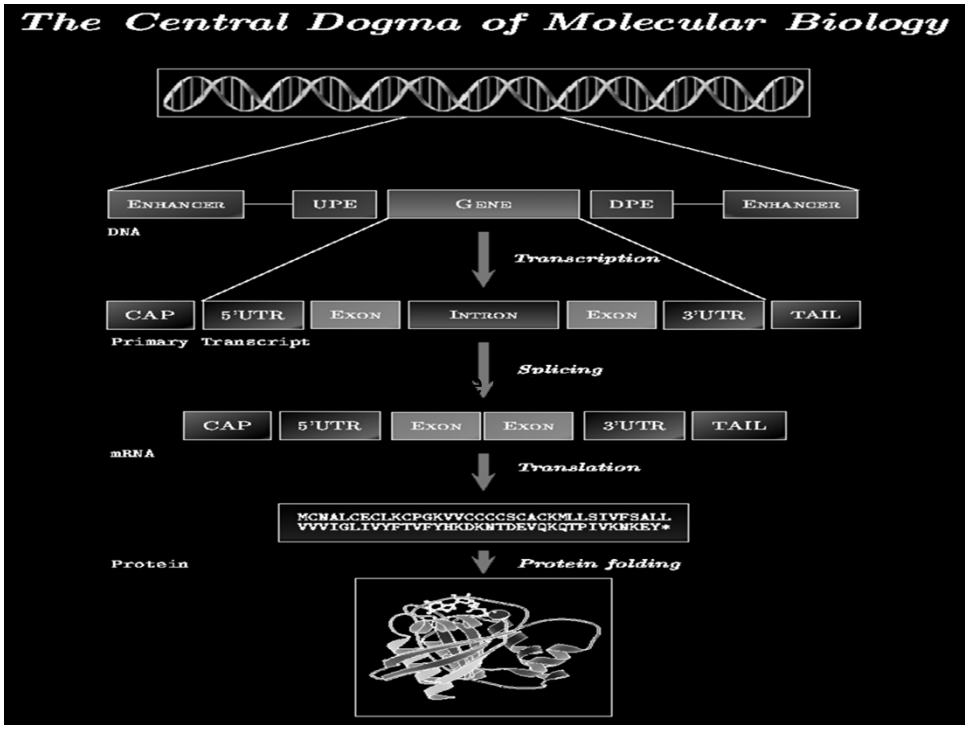
La molécula de ADN fue descubierta en 1951 por **James Watson**, **Francis Crick** y **Maurice Wilkins** empleando la técnica de difracción de los rayos X. En 1953, Watson (izquierda) y Francis Crick (derecha) describieron la estructura en doble hélice de la molécula de ADN como una especie de escalera de caracol con muchos escalones. En 1962 ambos recibieron el Premio Nobel de Medicina por su trabajo.



VNIVERSITAT DE VALENCIA

El Genoma Humano

CATGTTTGATTTTTTCAATTTTCGTCGCAGATTTTGTCGGTTTTGGGTACCTGGAACATTTCTCGACCA
ATAGCTCCGGAACTATTAGAAATAACCCCATAAAGTGTGCTATCGTTGGAAAGCTCTTTAAATTATCTA
TTTTTTCAGAAAAGATTGTTGTTCTCCGACTAATAGTTTTCGAGCAAATTGCTGCTAAATGCAAAAATT
GGAAATTTTATTGGTCAAATTAGCTAATAACTAATTTGTCCAGCAAGAACCACGTGGAGGTTTAGAGC
ATTCTTGCCAGGAAAAAGGATATAAGGATGAGGTAAACAATGATTACTTCAAATGAAGGGAATAAAGG
AATTATTGGTTTTGAAATAAAAAAATACAACACAGAAATAAAATGCACCTACCTTTATGGCTTCCACGT
TGGCTACTGTTTAGCAAAACAATAAAAAGAAATTAAAAACAACACCATTAAAAAATCTTAATTTGCATG
AAAATTACTACAGCGACATAATTTAAAGGTGTAAAAAACAATCAAGAATACTTATTATATTTAAGGATT
GCTCTAATTGGTCATGCAGTGTATTATCGTTAGAAGCTCTTTCAATTATCTATTTTTTTCACAAAAGATT
ATTGCTCTCCGACTAATAGTTTTTTTAAATATATAAAAAACTATTGGGAAGTTGGCAGTTGTCTCGATG
CTAATCGATTTCCCGGATCATTTTGCATAGGTAGATTAAAATAGTTCCACTTTTTCTAATAGTTTGTTCT
AGTGAATTCAGCGGCTAATTTTTGTTTATACGCTTTTCATTTATTTATAATTTATAATTTTGAATGTTTTT
GCTTAAAATTGTAATAATCAAGTCTTACATACGGATGTCCGGAACATTTATCTAGCAATAACAACGGAA
CTATTAGAGATAACCGACCGAACATAATCAACTGTGATACAGAACAATGGAATTATTGGTCAAAATATGGATAGTC
AATCCTCGTTGAGCTTGTAATTGAATTCCCCATGATTGGCATCAAAATTTGACCTTTTAATTTTTTACCATTTTTCGCAAATTTTGATGATGTTACCCCTTATCGAA
AATGCAAAATTGGACCCAAAAATTGATTTTTCAAAATCCTTCAAAAAATTGATTTTGTCAGTTTGTATTGGTAAAAAGCTGTAAAAAAAAAGCTATTGGTTCATCTTT
ATAAGCATTTTTGGCCAAGTTATGAAGAAAAGACCATTCGTAAATAGCGGATTTTTCCAAAAACCGATTTTCCGAATTTCCATAATACATCAATAGGTTTTTGGCCA
TAACTTTGAAATAATAGTCTGAATTCGGAATGTCATATATCGTTGAGCTCGTAATTAAATTTCCATTCAAACTGTATTCAAAATTGGAATTCTATTTTTTACCAATTTT
TTGCAAATTTTGATGATGTTACCCCTTACCGAAAATGCGAAAATCGACCCAAAATTTAATTTTTCAAAATCTTTCAAAAAGTGATAGGGATACTTAGTATTGGTAAA
AAGCTGTAAAAAAACAGCCATTGGTTTATCTTTATGACCATTTTTAGGCAAGTTACGATGTATAGATCTTTTGCAAAAAGCGGATTTTTCCAAAAACCGATTTTCCA
TAATAAAATAATCTGTTTTTTGGCCATAACTTTAAAAATAAGAGTCTGAATACGGAATGCCATATCTCGCTGAACTCGTAATTACATTTTCATTCAAACTGTATTCAA
AAATTGGAATTCTATTTTTTACCCATTTTTTGCAAATTTTGATGATGTTACCCCTTACCGAAAATGCGAAAATCGACCCAAAATTTAATTTTTCAAAATCTTTCAAAAA
GTGATAGGGATACTTAGTATTGGTAAAAAGCTGTAAAAAACAGTCATTGGTTTATTTTTATGACCATTTTTAGGCAAGTTACGATGTATAGATCTTTTGCAAAAAGCGGATTTTTCCAAA
AACCGATTTTCCATAATAAAATAATCTGTTTTTTGGCCATAACTTTAAAAATAAGAGTCTGAATATGGAATGCCATATCTCGTTGAGCTCGTAATTAAATTTCCATTCAAACTGTATTCAA
AAATTGGAATTCTATTTTTTCCCATTTTTTGCAAATTTTGATATGTTACCCCTTACCGAAAATGCGAAAATCGACCCAAAATTAAATTTAAAATCCCTCAAAAAGTGATTGGGATAATTAG
TATTGAATATCTCTTAAACTGTTATTCTTTTAACTATACGATCATTGTTAATCCAGTTTAGGCAAGTTACGATGTATAGATCTTTTGCAAAAAGCGGATTTTTCCAAAAACCGATTTTCCATAATAAAATAATCTGTTTTTTGG
CCATAACTTTAAAAATAAGA……



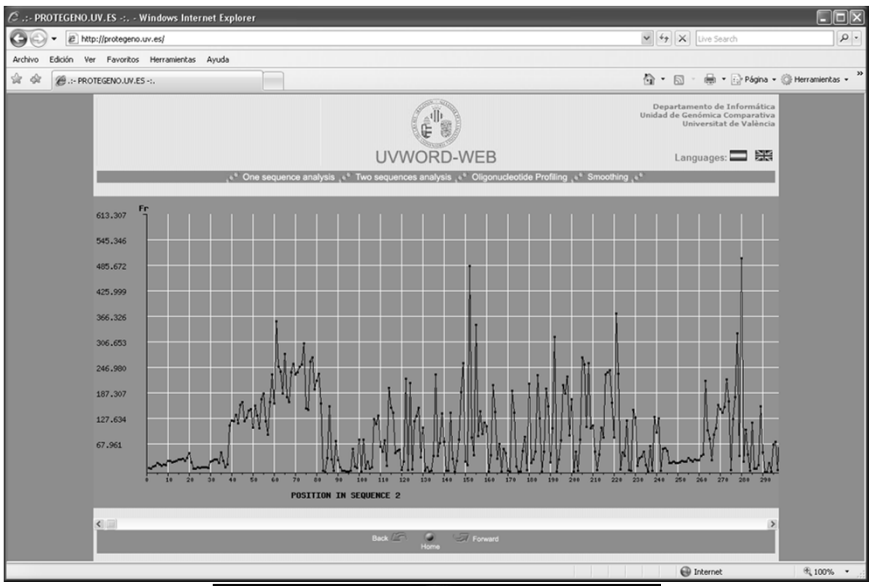
El Genoma Humano

PREGUNTA:
¿Cómo podemos saber que regiones de un gen o de un genoma son regiones codificantes y cuales no?.

RESPUESTA:

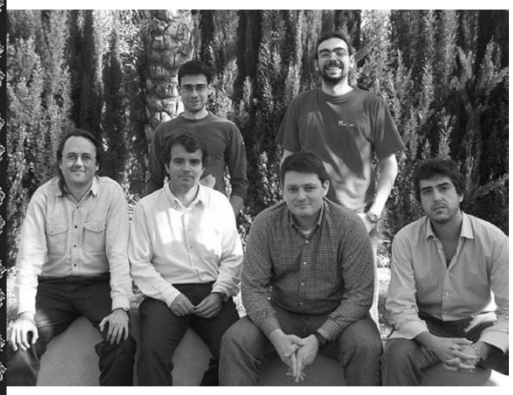
```
>Adh
TCACGTACTGTGGTCGTCCCTTGTTTATGGGCAGGCATCCCTCGTGCGTTGGACTGCTCG
TACATGTGGGCGAGGTTCCGTAAACGTCGGCATGTTGTCCACTGAGACAAACTTGTAAA
CCCGTTCCCGAACCAGCTGTATCAGAGATCCGTATTGTGTGGCCGTGGGGAGACCCTTCT
CGCTTAGCATCGAAAAGTAACCTGCGGGAAAAAGAAAAAATACAAATGTTAAAATTGTCC
TTGTATCTTATGTTGTATGCGTATCTCTTCTATTAAAGTGGGTTCATCTAACCATTATAC
ATTTCATAAATAAATAGATTACAATTGGGTCAAAATAAATGTTCAGTGAAGCTTCCCTT
CTCAAGGTCATAAAAGCATTTAAAAAAAATAGCACAAATCAATAATTAAAAACTAATTTT
GAAATCTCTTTGAACAAGACAGATATTTTGGTTCAGTCGCTGAACAAATCTGTTTACTGT
CTAAAATCTGAAAACCATTTTTCCGACAGCTGACAGCTTCGAACAGAATATAGTACACAA
TTTGCAGTCCAAAAATGAGTACAAAAAACAAAACAAATAATGACGACGCGACTGGGCAT
CTCTTAGTATTGAGATATATGTATTTAATTTTCTTAAAATAAAAGCATTTTTTGTCAATT
AAATGCAAAACCGACAAGTTTGATTGGAGGGTTTGTAAAAAATAAAATTCGAATGTAAAA
AGAGTTTCAGTTAGCGCAGGTGGATTTTACAGAAAAAAATGCAATGCAATTAAACATTAC
ATGTATCGATGAGTCCATTAATCATTTCATTTGGTTCAATTCGCGCCACTGAGCTTAAAT
TATAATGATACAATAAAAAAAATTGATGATAAAGAGAGACTA ...
```

El Genoma Humano



El Genoma Humano

GRUPO DE GENÓMICA DE LA UNIVERSIDAD DE VALENCIA



VNIVERSITAT DE VALÈNCIA

Departamento de Genética

Análisis de Patrones Globales de Evolución Genómica

TESIS DOCTORAL

Miguel Gallach Caballero

Valencia,2008

VNIVERSITAT DE VALÈNCIA

El Genoma Humano

Departamento de Informática
Unidad de Genómica Comparativa
Universitat de València
Instituto de Biomedicina (CSIC)

UVWORD-WEB

Languages:

Oligonucleotide Profiling | Oligo Smoothing | Oligo Counting | Singular Oligo Search

UVWORD-WEB is a web tool devised for fast screening of the oligonucleotides composing DNA sequences. The fast analysis of relative abundances of DNA-words (oligonucleotides) is very useful to detect specific oligonucleotides among discrete DNA sequences, chromosomes or even genomes.

UVWORD-WEB implements four kind of analysis based on the exhaustive computing of the oligonucleotides from 1 to 14 nucleotides-length present in any DNA sequence.

Oligonucleotide profiling tool computes for all oligonucleotide frequencies in a DNA sequence (Source). Later, oligos from a second DNA sequence (Target) are scanned to count their frequencies in Source. An additional parameter, window size, must be defined. Window size is the number of consecutive oligonucleotides the user wants to add together to obtain average counts. For window size = 1, no averages are calculated and only the frequencies for every one of the oligonucleotides are computed. For window size > 1, a third column with the average associated variances is added in the table.

In **Smoothing** option, unlike to the previous option, averages are calculated for sliding windows.

One sequence analysis counts all different oligonucleotides of a given length present in a DNA sequence. The user may choose the option "all" or "between" two defined frequencies (L_INF and L_SUP). The former returns a table in text format with all the oligonucleotides found in the sequence and their respective frequencies. The later returns a table with those oligos which frequencies are equal or higher than L_INF but also equal or lower than L_SUP.

In **Two sequences analysis** all different oligonucleotides are computed, and the analysis returns a table with all the oligos (and their respective frequencies) that are less or as frequent than a number F_INF in the first sequence and at the same time more or equally frequent than F_SUP in the second sequence.

For any question or suggestions contact with: **Vicente.Arnau@uv.es**.

http://protegeno.uv.es/

Back | Home | Forward

VNIVERSITAT DE VALÈNCIA

El Genoma Humano

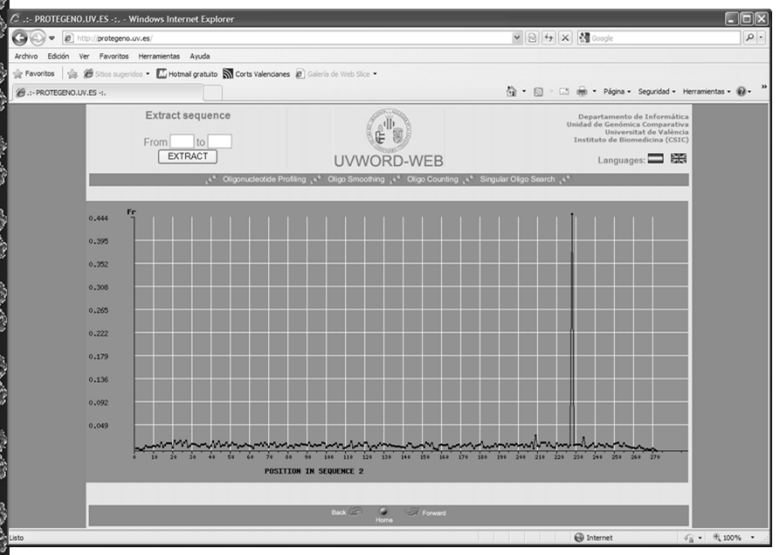
El Genoma Humano

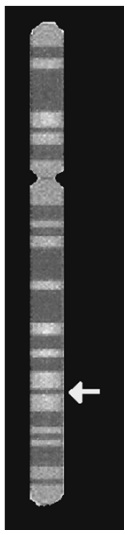
```
================================================
=                U V W O R D                   =
================================================
= V. Arnau; F. Ferri; I. Marin.   11 - V - 2005 =
================================================

 35058378 WORDS of 12 nucleotides in File human_chr22
 35449345 WORDS of 12 nucleotides in File human_chr21

 FINF = 2    FSUP = 60


================================================
   WORD          Frec_1          Frec_2
------------------------------------------------

AAATGTCCACTT          1              98
AACTCTGTGAGT          2              62
AATGGATTGGAA          2              61
AATGTCCACTTG          1              82
ATATCCACTTGC          1              72
ATGGAATGCAAT          2             151
ATGGACTGGAAT          2              82
ATGTCCACTTGC          1              80
CAAATGTCCACT          1              79
CAATGGAATGGA          2             190
CATTGTTTTGGG          2              75
CCCATTGTTTTG          2              60
CTTGAATGGAAT          2              70
GAAGCGCATTCA          0              62
GAATGGAATGTA          2              76
GAATGGACTGGA          1              85
GATTGGAATGGA          2              74
GCAATGGAATGG          2             144
GGAATGGACTCG          0             261
GGAATGGACTGG          1              71
GTCCACTTGCAG          2              81
GTGGAATGGAAT          2             300
GTGGAATGGAGT          1             114
GTGGAGTGGAAT          1             163
TCTGAGAAACTT          2              65
TGAGTGGTGAGT          2             142
TGAGTTGAATGC          1              60
TGCAATGGAATG          2             155
```

Listo

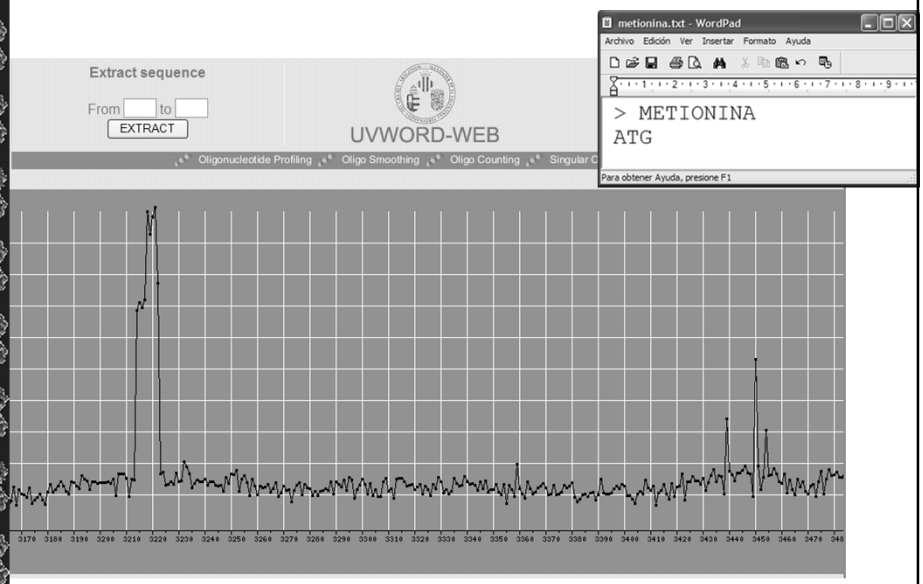
Internet | Modo protegido: activado

El Genoma Humano



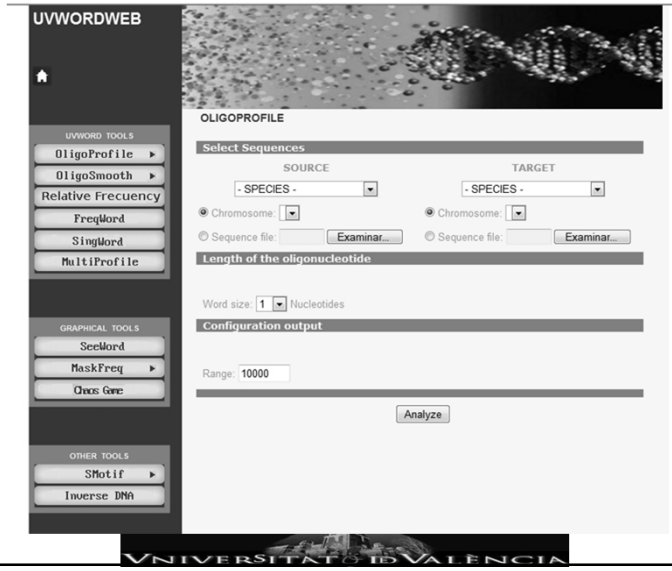


El Genoma Humano

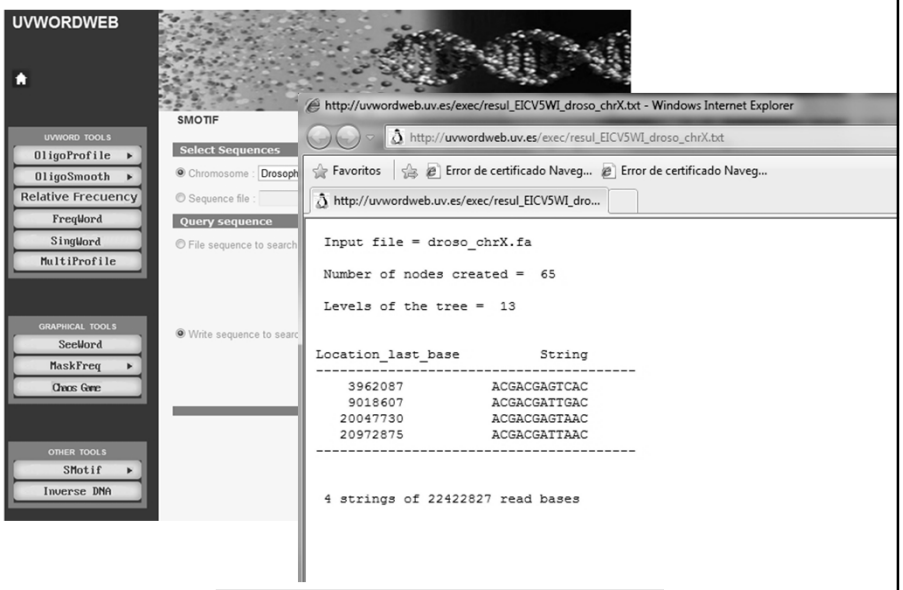


El Genoma Humano

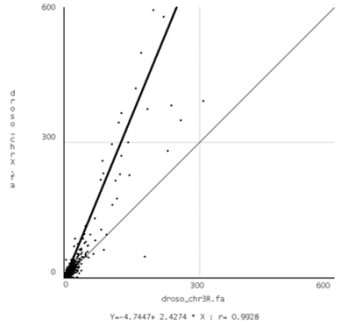
http://uvwordweb.uv.es/

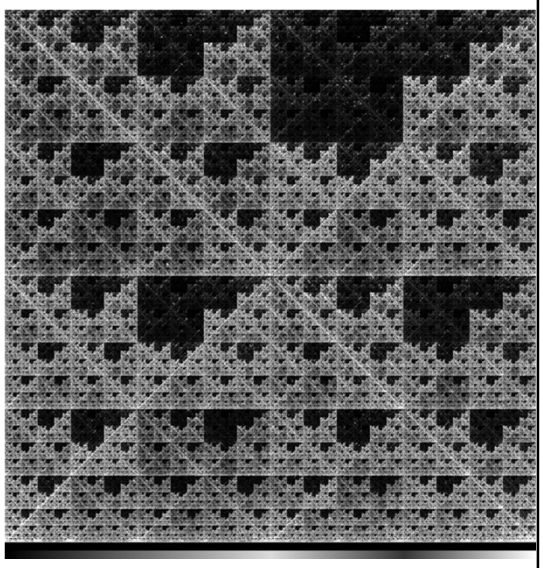


El Genoma Humano

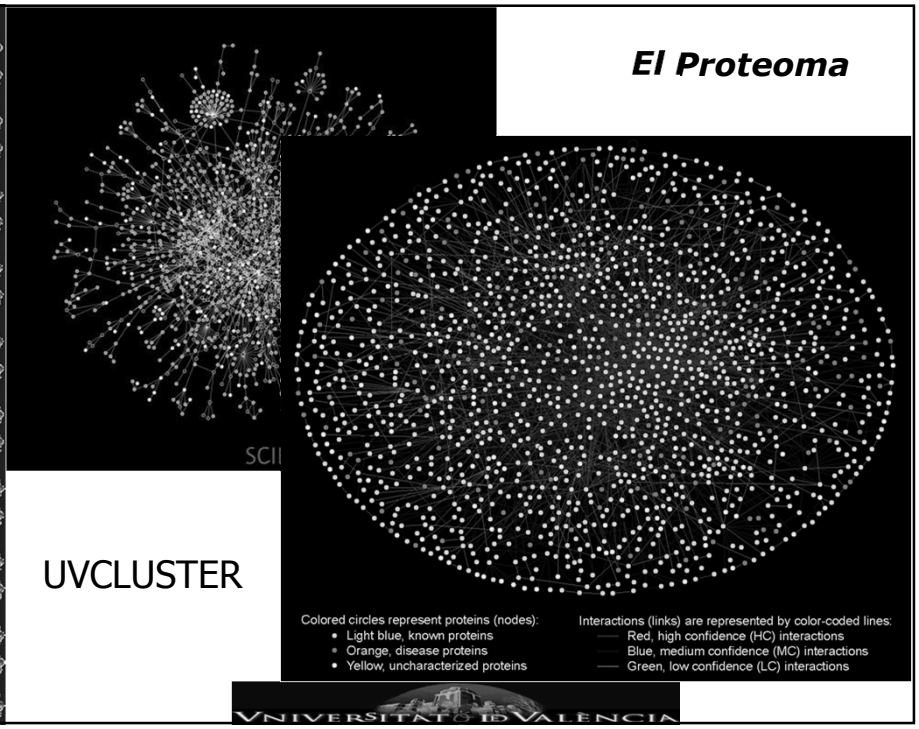


El Genoma Humano





El Proteoma



UVCLUSTER

El Proteoma

G R A F O D E I N T E R A C C I O N E S

- Leemos de la base de datos un conjunto de interacciones directas entre proteínas y creamos el grafo.

Bni1 ↔ Act1

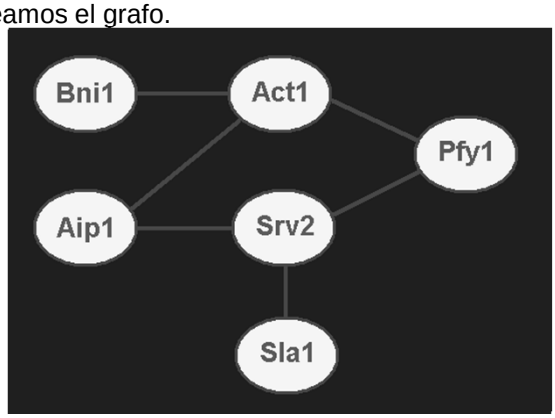
Aip1 ↔ Srv2

Aip1 ↔ Act1

Pfy1 ↔ Act1

Pfy1 ↔ Srv2

Srv2 ↔ Sla1

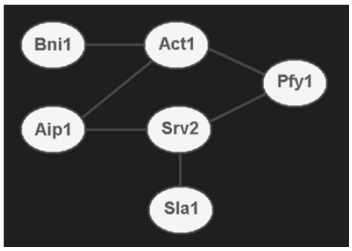


VNIVERSITAT DE VALÈNCIA

El Proteoma

UVCLUSTER: ANÁLISIS DE INTERACCIONES ENTRE PROTEÍNAS

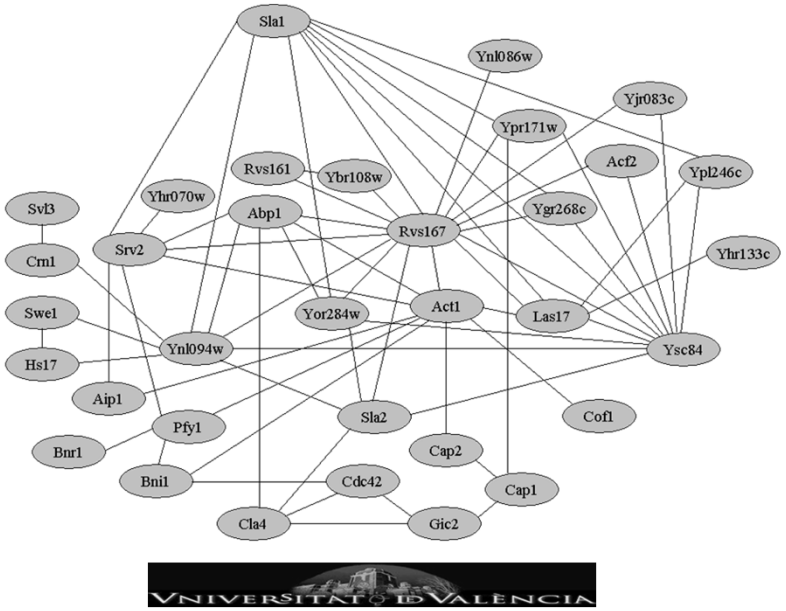
Convertimos Grafo de Interacciones en Tabla de Distancias entre proteínas.



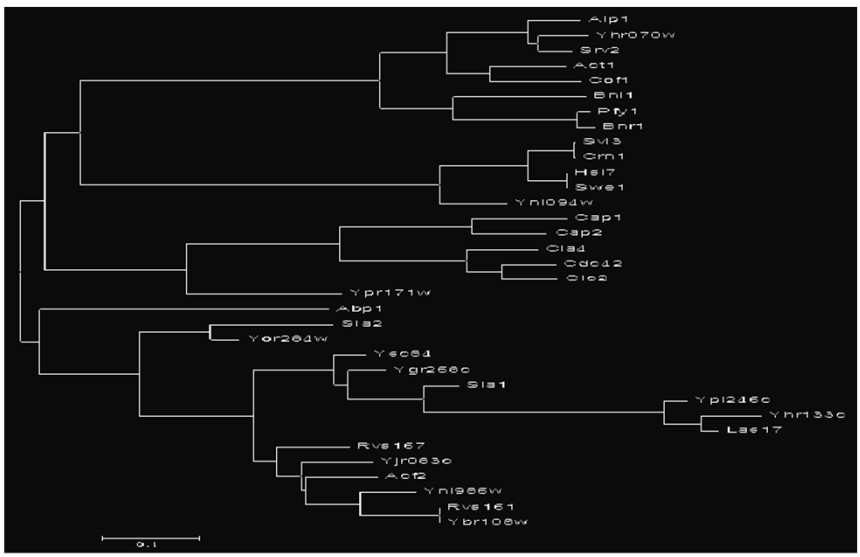
	Bni1	Act1	Aip1	Srv2	Pfy1	Sla1
Bni1	0	1	2	3	2	4
Act1	1	0	1	2	1	3
Aip1	2	1	0	1	2	2
Srv2	3	2	1	0	1	1
Pfy1	2	1	2	1	0	2
Sla1	4	3	2	1	2	0

VNIVERSITAT DE VALÈNCIA

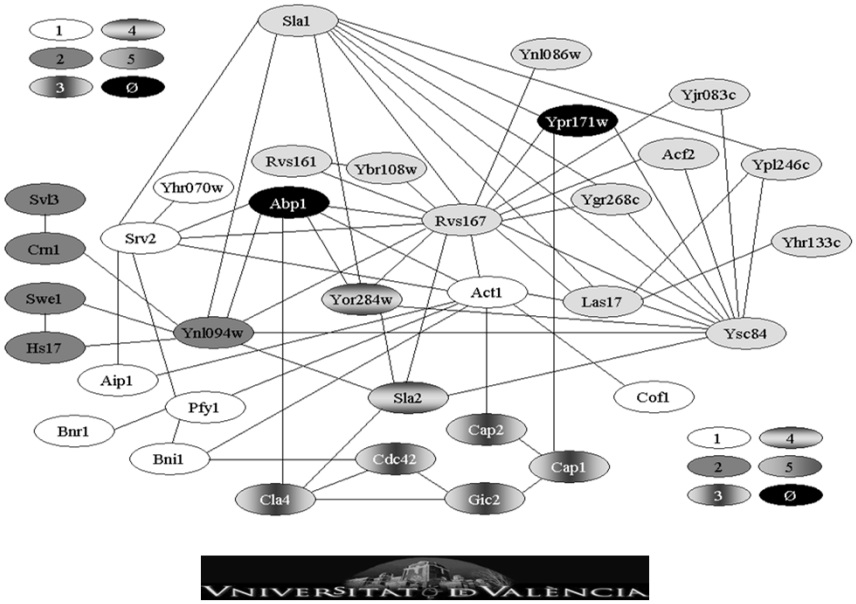
El Proteoma



El Proteoma

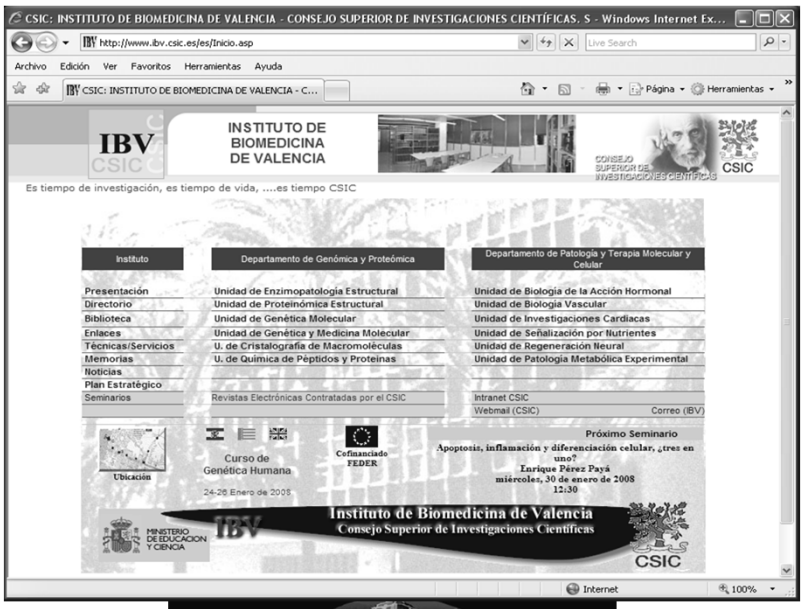


El Proteoma



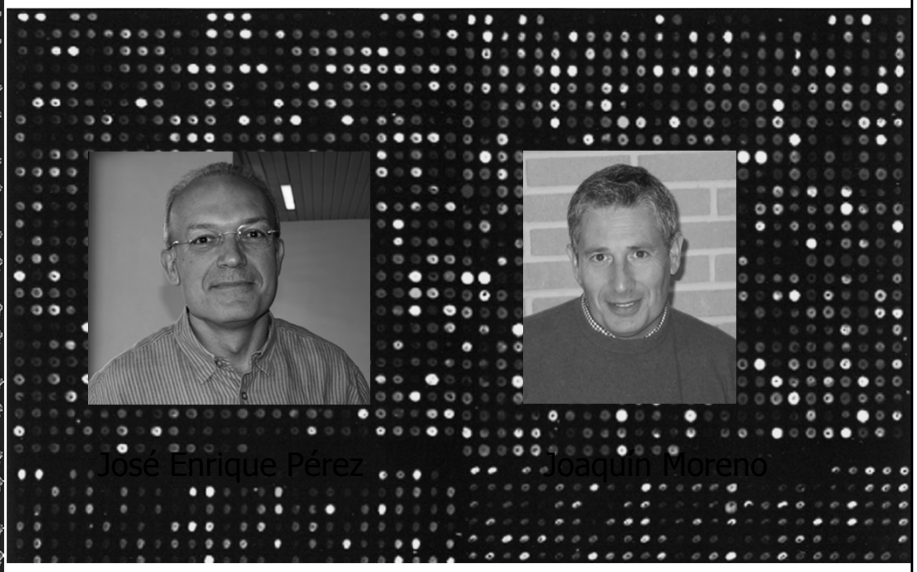
VNIVERSITAT DE VALÈNCIA

El Genoma Humano



VNIVERSITAT DE VALÈNCIA

BIOINFORMÁTICA



José Enrique Pérez Joaquín Moreno

VNIVERSITAT DE VALÈNCIA

El Genoma Humano

VNIVERSITAT DE VALÈNCIA

El Genoma Humano

BIOINFORMÁTICA

Vicente Arnau Llombart

http://www.uv.es/~varnau/bioinformatica.pdf

E-mail: Vicente.Arnau@uv.es