



BIOINFORMÁTICA

Vicente Arnau Llombart

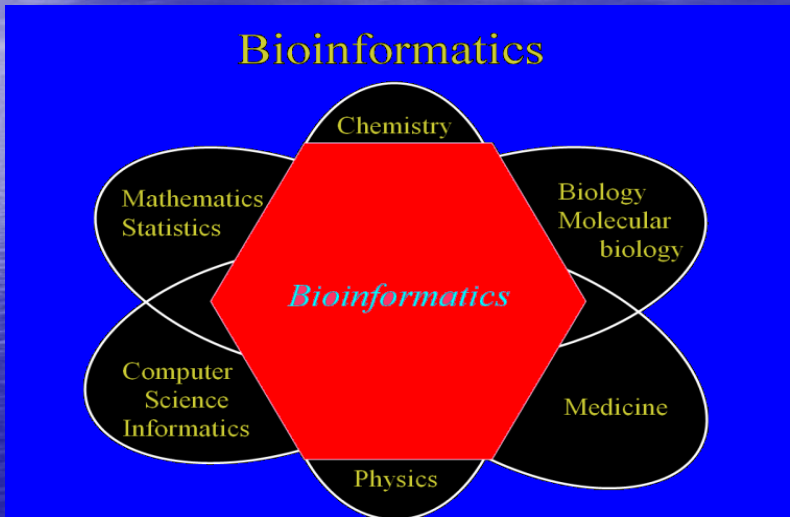
<http://www.uv.es/~varnau/bioinformatica.pdf>

E-mail: Vicente.Arnau@uv.es



¿Que es la BIOINFORMÁTICA?

BIOLOGIA MOLECULAR ↔ INFORMÁTICA



Bioinformatics

Chemistry

Mathematics Statistics


Computer Science Informatics

Physics

Biology Molecular biology

Medicine

Bioinformatics



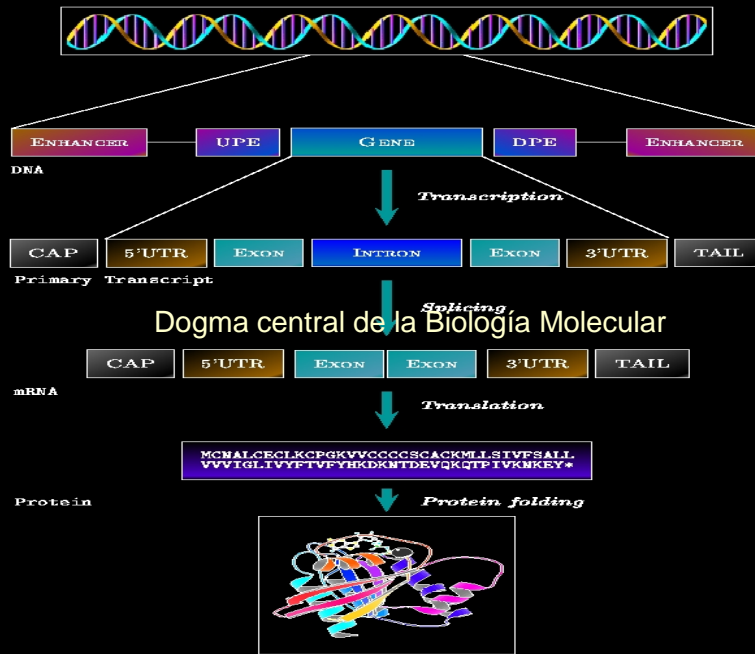
UNIVERSITAT DE VALÈNCIA

El Genoma Humano

CATGTTTGATTTTTCAATTTTCGTCGCAGATTTGTCGGTTTTGGGTACCTGGAACATTTCTCGACCA
 ATAGTCCGGAECTATTAGAAATAACCCATAAAGTGTGCTATCGTTGGAAGCTCTTTAAATTATCTA
 TTTTTTCAGAAAAGATTGTTGTTCTCCGACTAATAGTTTTCGAGCAAATGCTGCTAAATGCAAAAATT
 GGAAATTTTATGGTCAAATAGCTAATAACTAATTTGTCCAGCAAGAACCACGTGGAGGTTTAGAGC
 ATTCTTGCCAGGAAAAAGGATATAAGGATGAGGTAAACAATGATTACTTCAAATGAAGGGAATAAAGG
 AATTATTGGTTTTGAAATAAAAAAATAACAACACAGAAATAAAATGCACCTACCTTTATGGCTCCACGT
 TGGCTACTGTTAGCAAAAACAATAAAAAAGAAATAAAAACAACACCATAAAAAATCTTAATTTGCATG
 AAAATTACTACAGCGACATAATTTAAAGGTGTAAAAACAATCAAGAATACTTATATATTTAAGGATT
 GCTCTAATTTGGTCATGCAGTGTATTATCGTTAGAAGCTCTTCAATTATCTATTTTTTCCAAAAAGATT
 ATTGCTCTCCGACTAATAGTTTTTTAAATATATAAAAAACTATTGGGAAGTTGGCAGTTGCTCGATG
 CTAATCGATTTCCCGGATCATTTTGCATAGGTAGATTAAATAGTCCACTTTTTCTAATAGTTTGTCT
 AGTGAATCAGCGGCTAATTTTTGTTATACGCTTTTCATTTATTTATAATTTATAATTTGAATGTTTT
 GCTTAAATTTGAATAATCAAGTCTTACATACGGATGTCGGGAACATTTATCTAGCAATAACAACGGAA
 CTATTAGAGATAACCGACCGAACATAATCAACTGTGATACAGAACAATGGAATTATTGGTCAAAAATGGATGTC
 AATCCTCGTGGAGCTTGAATTGAATCCCATGATTGGCATCAAAATTTGACCTTTAATTTTTACCATTTTCGCAAAATTTGATGATTTACCCCTTATCGAA
 AATGCAAAATTTGGACCCAAAAATTTGATTTTCAAAATCTTCAAAAAATTTGATTTTGTCAAGTTTGTATTGGTAAAAAGCTGTAAAAAAAAGCTATTGGTTCATCTTT
 ATAAGCATTTTTGGCCAAAGTTATGAAGAAAAGACCATTCGTAATAGCGGATTTTCCAAAAACGATTTCCGAAATTTCCATAACATCAATAGGTTTTGGCCAA
 TAACCTTGAATTAATAGTCTGAATTCGGATGTCATATATCGTTGAGCTGTAATTAATTTGCATTGAAAGCTGATTTCAAAATTTGGATTCTAATTTTTCCAAATTTT
 TTGCAAAATTTGATGATGTTACCCCTTACCGAAAATGCGAAAATCGACCAAAATTTAATTTTCAAAATCTTTCAAAAAGTATAGGATACTAGTATTGCGTAAA
 AAGCTGTAAAAAACAGCCATTTGGTTATCTTTATGACCATTTTAGGCAAGTTACGATGTATAGATCTTTTCCAAAAAGCGGATTTTCCAAAAACGATTTTCCA
 TAATAAAATAATCTGTTTTTGGCCATAACCTTAAAAAAGAGTCTGAATACGGAATGGCCATATCTCGTGAACCTGTAATACATTTTCAATTCAACTGATTCAA
 AAATTTGGAATCTATTTTTACCCATTTTTGCAAAATTTGATGATGTTACCCCTTACCGAAAATGCGAAAATCGACCCAAAAATTTAATTTTTCAAAATCTTCAAAAA
 GTGATAGGACTAGTATTGGTAAAAAGCTGAAAAACAGCTATTGGTTTTTTTATGACCATTTTAGGCAAGTTACGATGTATAGATCTTTGCAAAAAGCGGATTTTCCAAA
 AACCGATTTCCATAATAAAATAATCTGTTTTTGGCCATAACTTAAAAAAGAGTCTGAATACGGAATGGCCATATCTCGTGAACCTGTAATACATTTTCAATTCAACTGATTCAA
 AAATTTGGAATCTATTTTTTCCATTTTTTCCAAATTTTGAATGTTACCCCTTACCGAAAATGCGAAAATCGACCCAAAAATTTAATTTAATAACCCCTCAAAAAGTATTCAA
 TATTGATATGTTAAACTGTTATCTTTAACTATACGATCATTTGTTAATCCAGTTTGGGAAAGTACGATGTATAGATCTTTTCCAAAAGCGGATTTTCCAAAAGCGGATTTTCCAAA
 CCATACTTAAAAATAGA.....



The Central Dogma of Molecular Biology



El Genoma Humano

PREGUNTA:

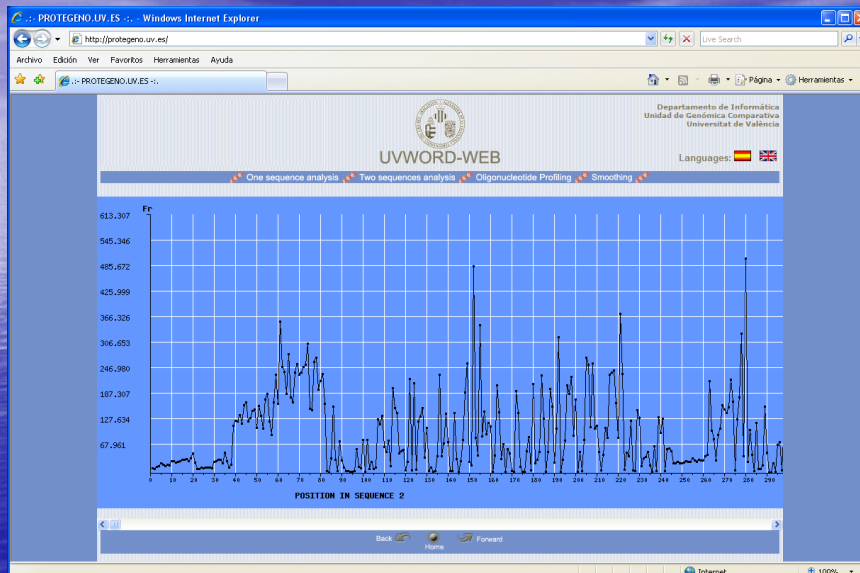
¿Cómo podemos saber que regiones de un gen o de un genoma son regiones codificantes y cuales no?.

RESPUESTA:

```
> Adh
TCAOGTACTGTGGTCGTCOCCTGTTTATGGGCAGGCATCCCTCGTGCCTGGACTGCTG
TACATGTGGCGGAGGTTCCGTAACGCTGGCATGTGTCCACTGAGACAAACTGTAAAA
COCGTTCCOGAAACCAAGCTGTATCAGAGATCOGTA TTGTGTGGCCTGGGGAGACOCCTTCT
CGCTTAGCATOGAAAAGTAACCTGGGGAAAAGAAAATAACAAATGTAAAAATTGTCC
TTGTACTTTAAGTTGTATGCGTATCTCTCTATAFAAAGTGGGTTTCATCTAACCAATTATAC
ATTTTCATAAAAATAAGATTACAATTTGGGTCAAAAATAAATGTTCACTGGAAGCTTCCOCTT
CTCAAGGTCATAAAGCATTFAAAAATAAGCACAAAATCAATAAATAAACAATAATTT
GAAATCTCTTTGAACAAGACAGATATTTTGGTTCAGTCGCTGAACAAAATCTGTTTACTGT
CTAAAATCTGAAAACCA TTTTCCGACAGCTGACAGCTTCGAAACAGAATATAGTACACAA
TTTTGCAGTCCAAAATGAGTACAAAACAAAACAAATAATGGAOGAOGOGACTGGGCAT
CTCTTAGTATTGAGATATA TGTA TTTAA TTTCTTAAAATAAAGCATT TTTTGTCAATT
AAAATGCAAAAACGACAAAGTTTGA TTGGAAGGTTTGTAAAATAAATAATGAAATGTA AAA
AGAGTTTCAGTTAGCCAGGTTGGATTTACAGAAAATAAGCAATGCAAATTAACATTAC
ATGATTCGATGAGTCCATTAA TCATTTTCATTTGGTTCAATFOGOGCCACTGAGCTTAAAT
TAT AATGATACAAATAAAAATAATGATGATAAAGAGACTA . . .
```

UNIVERSITAT DE VALÈNCIA

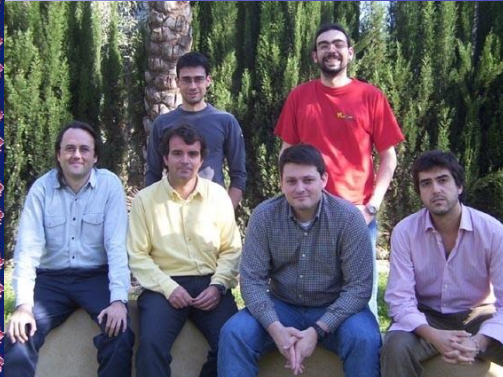
El Genoma Humano



UNIVERSITAT DE VALÈNCIA

El Genoma Humano

GRUPO DE GENÓMICA DE LA UNIVERSIDAD DE VALENCIA



VNIVERSITAT ID VALENCIA

Departamento de Genética



Análisis de Patrones Globales de Evolución Genómica

TESIS DOCTORAL
Miguel Gallach Caballero
Valencia, 2008



El Genoma Humano



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

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.Amas@uv.es.

<http://protegeno.uv.es/>

[Back](#) [Home](#) [Forward](#)



El Genoma Humano



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

Languages:

[Oligonucleotide Profiling](#) [Oligo Smoothing](#) [Oligo Counting](#) [Singular Oligo Search](#)

SEQUENCE 1: Homo sapiens
 Chromosome 1: Sequence file 1:
 Chromosome 22:

SEQUENCE 2: Homo sapiens
 Chromosome 1: Sequence file 2:
 Chromosome 21:

Word size: 12 nucleotides
 FREQUENCIES:
 Frequency LESS OR EQUAL to 2 in SEQUENCE 1 and GREATER OR EQUAL to 60 in SEQUENCE 2

Use cache

Singular Oligo Search

The idea of this routine is to find out sequences which are enriched in one sequence and rare in another one. The analysis returns a table with all the oligos (and their respective frequencies) which are present at most F_INF times in the first sequence and at least F_SUP times in the second sequence.

[Back](#) [Home](#) [Forward](#)



El Genoma Humano

```

-----
      W O R D
-----
- V. Arnau: F. Ferri: I. Marin. 11 - V - 2005 -
35058378 WORDS of 12 nucleotides in File human_chs22
35449345 WORDS of 12 nucleotides in File human_chs21
FINF = 2   FSUP = 60
-----

```

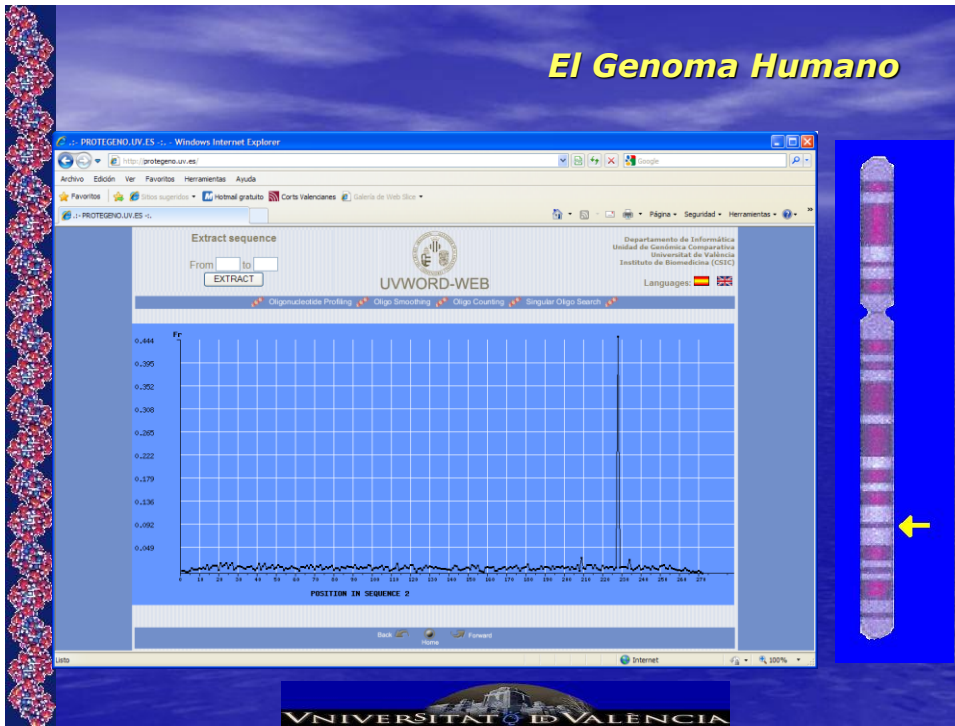
WORD	Frec_1	Frec_2
AAATGCCACTT	1	98
AACCTCTGAGT	2	62
AATGATGGAAA	2	61
AATGCCACTTG	1	82
ATATCCACTGC	1	72
ATGGAATGCAAT	2	151
ATGGACTGGAT	2	82
ATGCCACTTGC	1	80
CAATCTGCCT	1	79
CAATGGAATGGA	2	190
CAATGTTTGGG	2	75
CCCATGTTTTG	2	60
CTTGAATGGAT	2	70
GAAAGCCATTCA	0	62
GAAAGGAATGA	2	76
GAAAGGACTGGA	1	85
GAAAGGAATGGA	2	74
GCAATGGATGG	2	144
GGATGGACTTGG	0	261
GGATGGACTGG	1	71
GTCCACTTGCAG	2	81
GTGGAATGGAAT	2	300
GTGGAATGGAGT	1	114
GTGGATGGAAAT	1	163
TCTGGAACTT	2	65
TGAGTGGTGGT	2	142
TGAGTGAATGC	1	60
TGCAATGGAATG	2	155

Lista

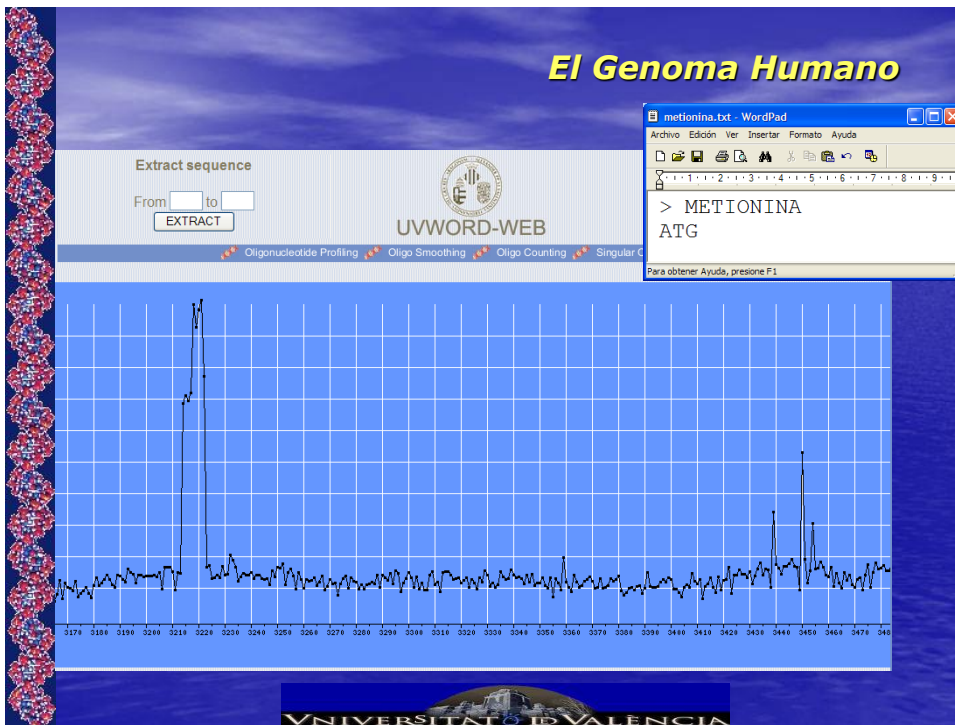
Internet | Modo protegido: activado



El Genoma Humano



El Genoma Humano



El Genoma Humano

<http://uvwordweb.uv.es/>

The screenshot shows the UWORDWEB website interface. On the left is a navigation menu with categories: UWORD TOOLS (OligoProfile, OligoSmooth, Relative Frequency, FreqWord, SingWord, MultiProfile), GRAPHICAL TOOLS (SeqWord, MaskFreq, Choc Gene), and OTHER TOOLS (SMotif, Inverse DNA). The main content area is titled 'OLIGOPROFILE' and contains a 'Select Sequences' section with 'SOURCE' and 'TARGET' dropdowns for '- SPECIES -'. Below are radio buttons for 'Chromosome:' and 'Sequence file:' with 'Examinar...' buttons. A 'Length of the oligonucleotide' section has a 'Word size:' dropdown set to '1' and 'Nucleotides'. A 'Configuration output' section has a 'Range:' input set to '10000' and an 'Analyze' button. A decorative DNA helix border is on the left, and the 'UNIVERSITAT DE VALÈNCIA' logo is at the bottom.

El Genoma Humano

This screenshot shows the UWORDWEB website with the 'SMOTIF' tool selected. The browser window displays the following output:

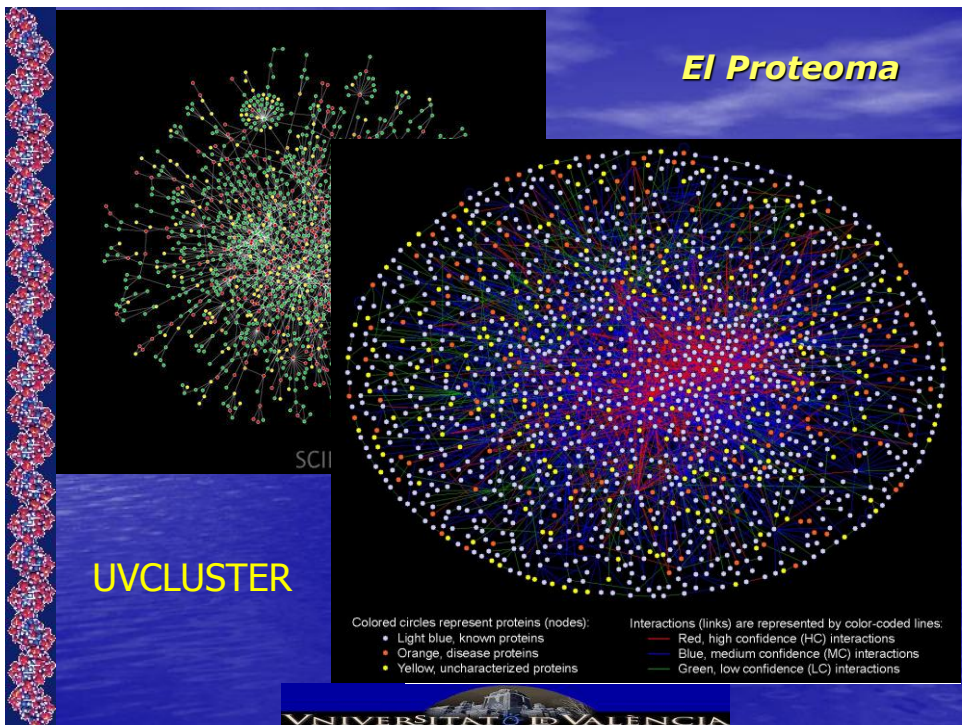
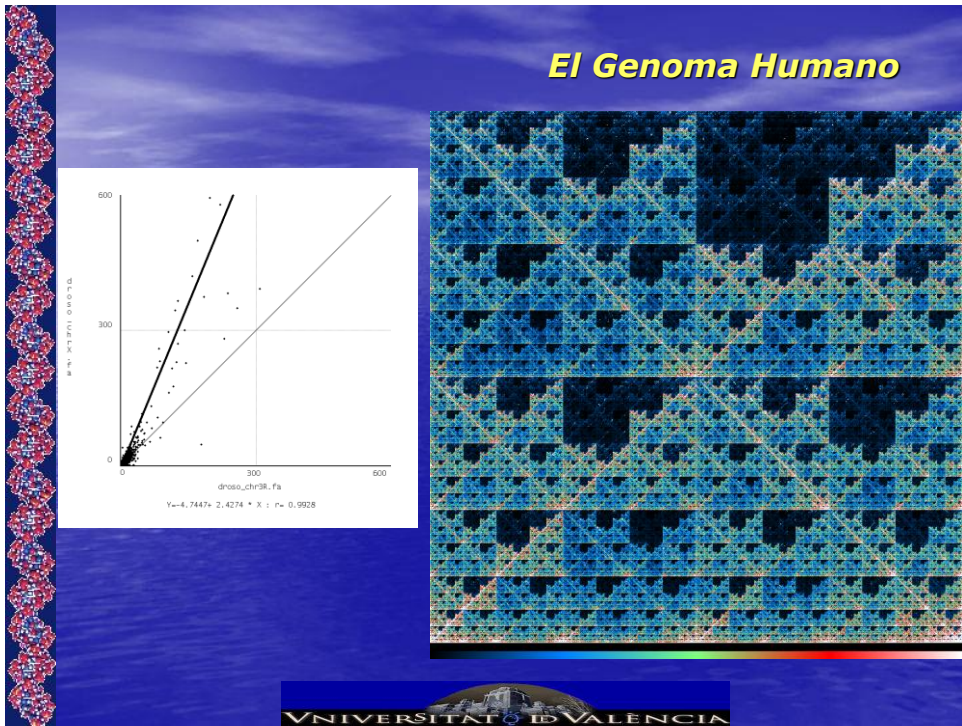
```
http://uvwordweb.uv.es/exec/resul_EICV5WI_droso_chrX.txt - Windows Internet Explorer
http://uvwordweb.uv.es/exec/resul_EICV5WI_droso_chrX.txt
Favoritos Error de certificado Naveg... Error de certificado Naveg...
http://uvwordweb.uv.es/exec/resul_EICV5WI_dro...

Input file = droso_chrX.fa
Number of nodes created = 65
Levels of the tree = 13

Location_last_base      String
-----
3962087      ACGACGAGTAC
9018607      ACGACGATTAC
20047730     ACGACGAGTAC
20972875     ACGACGATTAC
-----

4 strings of 22422827 read bases
```

The website interface on the left is similar to the first screenshot, but with 'SMOTIF' selected in the 'OTHER TOOLS' menu. The 'UNIVERSITAT DE VALÈNCIA' logo is at the bottom.



GRAFO DE INTERACCIONES

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

Bni1 ↔ Act1

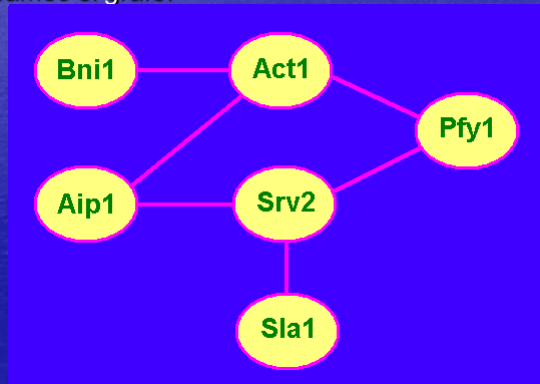
Aip1 ↔ Srv2

Aip1 ↔ Act1

Pfy1 ↔ Act1

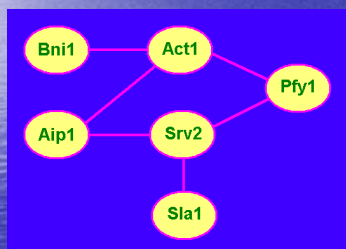
Pfy1 ↔ Srv2

Srv2 ↔ Sla1



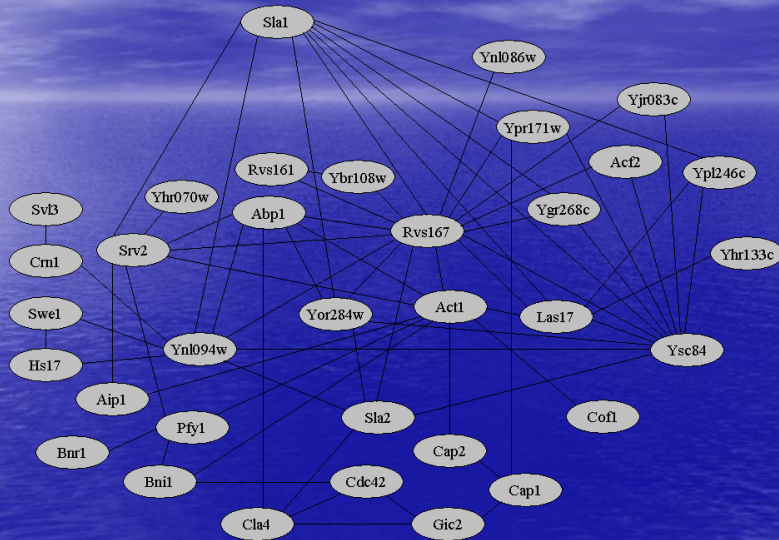
UVCLUSTER: ANÁLISIS DE INTERACCIONES ENTRE PROTEÍNAS

Convertimos Grafo de Interacciones en Tabla de Distancias entre proteínas.



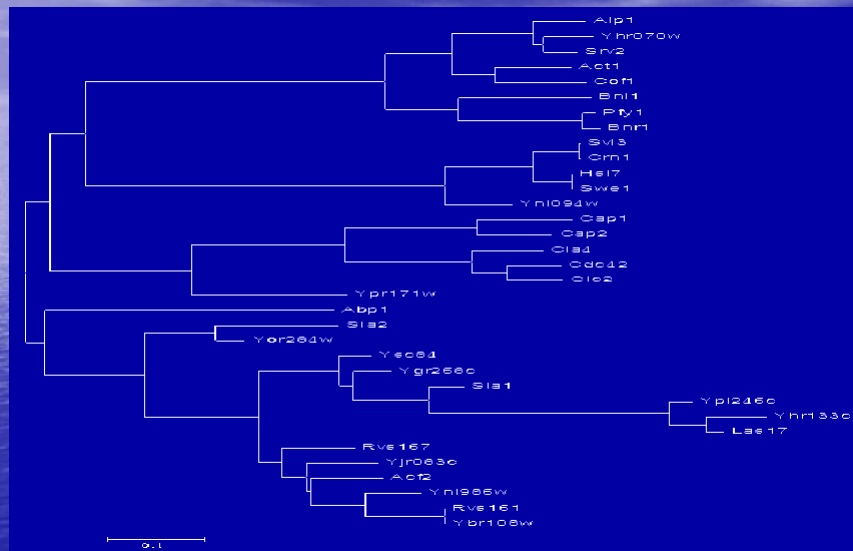
	Bni1	Act1	Aip1	Srv2	Pfy1	Sla1
Bni1	0	1	2	3	2	4
Act1	1	0	1	2	1	3
Aip1	2	1	0	1	2	2
Srv2	3	2	1	0	1	1
Pfy1	2	1	2	1	0	2
Sla1	4	3	2	1	2	0

El Proteoma

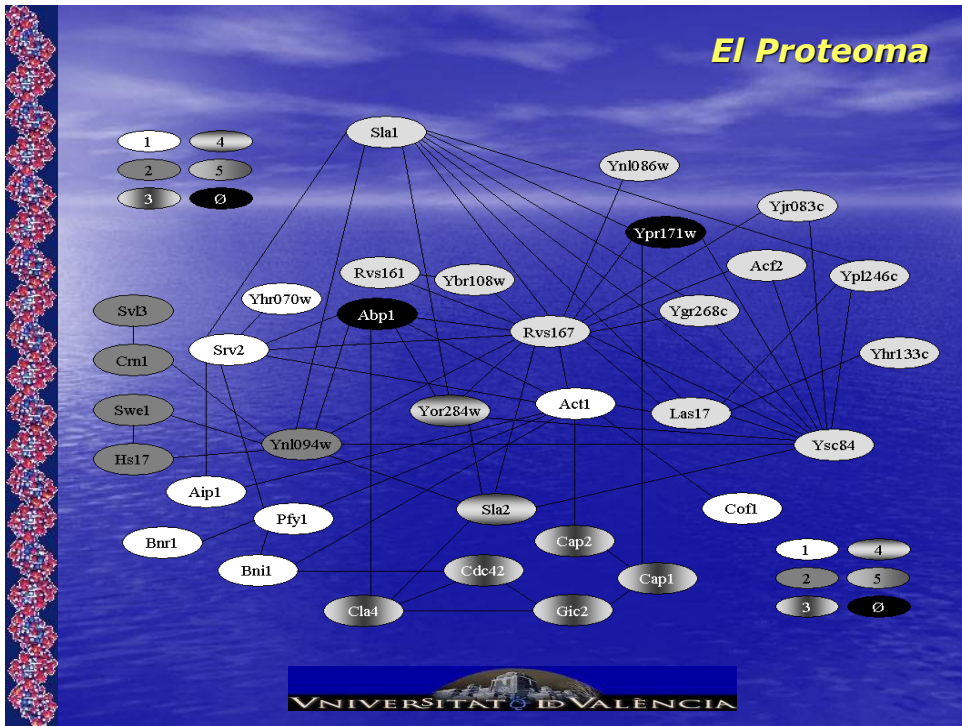


UNIVERSITAT ID VALENCIA

El Proteoma



UNIVERSITAT ID VALENCIA



El Genoma Humano

UNIVERSITAT DE VALÈNCIA

BIOINFORMÁTICA



José Enrique Pérez



Joaquín Moreno

UNIVERSITAT DE VALÈNCIA

BIOINFORMÁTICA



```

P L N I E V P K I S
P L N I E V P K I S
5'-CCTCTCAACATTGAGGTCCCAAAATCA
3'-GGAGAGTTGTAACCCAGGGGTTTAGT
3'-GGAGCGTTGTAACCCAGGGGTTTAGT
3'-GTTGTAACCCAGGGGTTTAGT
3'-AACTCCAGGGGTTTAGT
5'-ctccaggggttttagt
3'-CCAGGGGTTTAGT
5'-ggggtttagt
3'-TTTTGGT
3'-TTTAGT
3'-GT
5'-
3'-
5'-
GGTCCGATGAGGTTGAGGTCGTCACAGAAAGAACTACAAGAAG-5'
5'-tcggagtaagagttgaaaagtcgtcacaggaagaactacaagaagtc-3'
3'-GAGTAAGAGTAGAAAAGTCGTACAGGAAAGAACTACAAGAAGTCACTC-5'
5'-agagttgaaaagtcgtcacaggaagaactacaagaagtcactccccgg-3'
3'-GTTGAAAAGTCGTACAGGAAAGAACTACAAGAAGTCACTCCCCGGAAT-5'

```

JOAQUIN DOPAZO: Secuenciación masiva de ADN

UNIVERSITAT DE VALÈNCIA

El Genoma Humano

Vicente Arnau Llobart

- [Datos Personales.](#)
- [Docencia.](#)
- [Investigación.](#)
- [Más Personal.](#)

Visitas= 0.445.49

UNIVERSITAT DE VALÈNCIA

El Genoma Humano

Investigación

- [Participación en Proyectos.](#)
- [Publicaciones y congresos.](#)
- [Patentes Software](#)
- [Aplicaciones WEB diseñadas.](#)
- [Tesis y Tesinas dirigidas.](#)
- [Cursos impartidos.](#)
- [Premios.](#)

Grupo de trabajo

[Unidad de Genómica Comparativa.](#)

[\[Home\]](#)

UNIVERSITAT DE VALÈNCIA



BIOINFORMÁTICA

Vicente Arnau Llombart

<http://www.uv.es/~varna/bioinformatica.pdf>

E-mail: Vicente.Arnau@uv.es

