

Búsqueda de Motivos Biológicos en Secuencias de ADN de origen Ambiental

Luis Garreta¹, Patricia Vélez², Pedro Moreno³, e Irene Tischer⁴

Centro de Excelencia en Genómica Y Bioinformática De Ambientes Extremos -
GeBiX

Grupo de Bioinformática y Biocomputación
Universidad del Valle -
Cali-Colombia



Agenda

① Introducción

② El Enfoque

③ Resultados

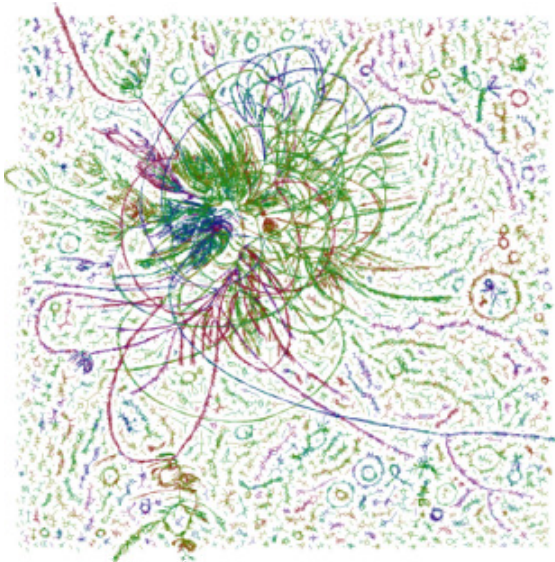
Introducción

GEBIX

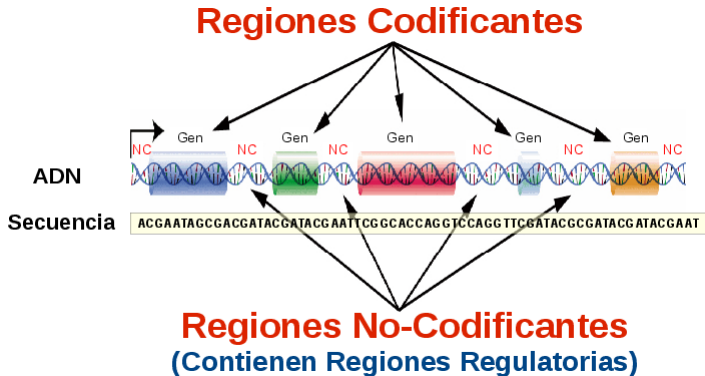
Centro Colombiano de Genómica y Bioinformática de Ambientes Extremos



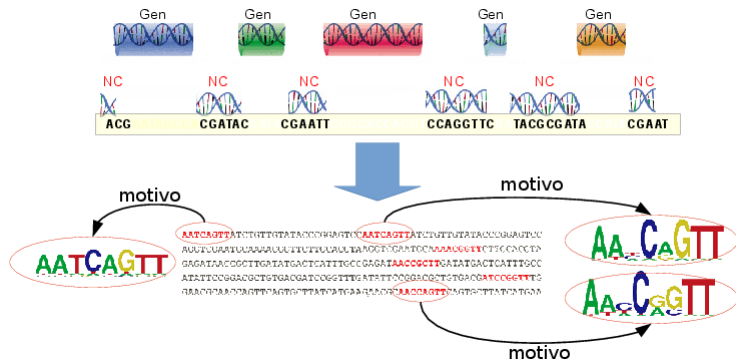
Metagenómas



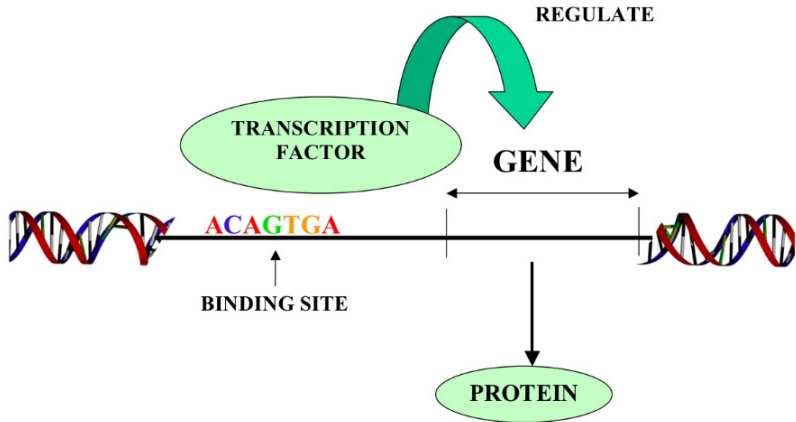
Regiones No-Codificantes



Extracción de las Regiones No-Codificantes



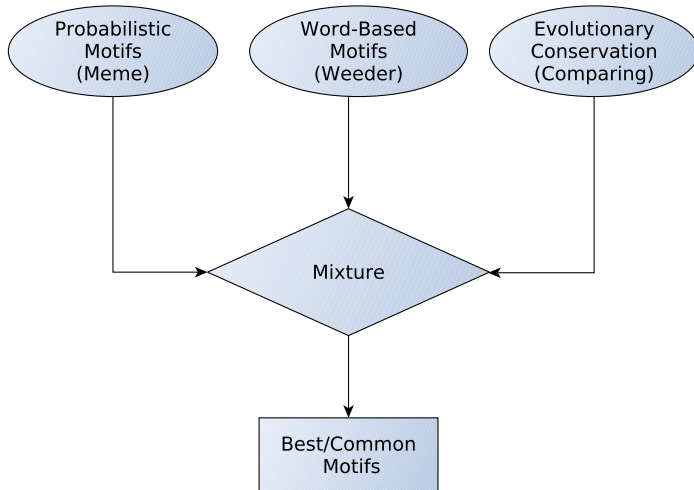
Sitios de Enlace de Factores de Transcripción



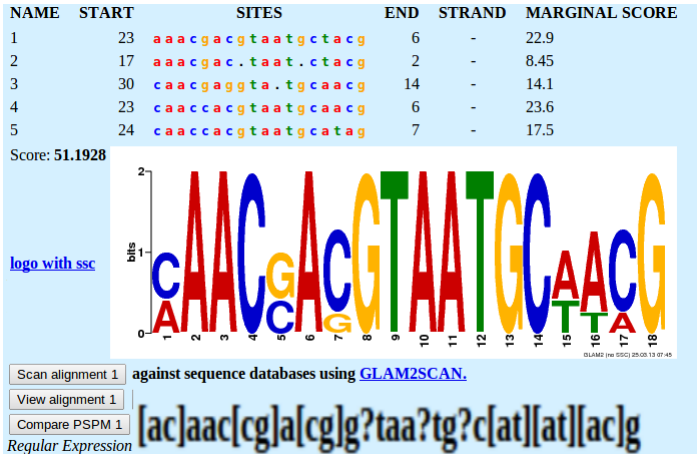
Enfoque

Múltiples Fuentes:

Mezcla de Expertos



Word-based Method



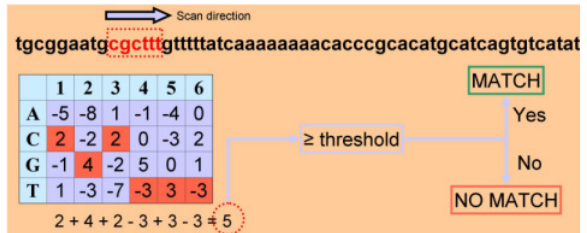
Probabilistic Method

a HEM13 CCCATTGTTCTC
 HEM13 TTTCTGGTTCTC
 HEM13 TCAATTGTTTAG
 ANB1 CTCATTGTTGTC
 ANB1 TCCATTGTTCTC
 ANB1 CCTATTGTTCTC
 ANB1 TCCATTGTTCTG
 ROX1 CCAATTGTTTTC

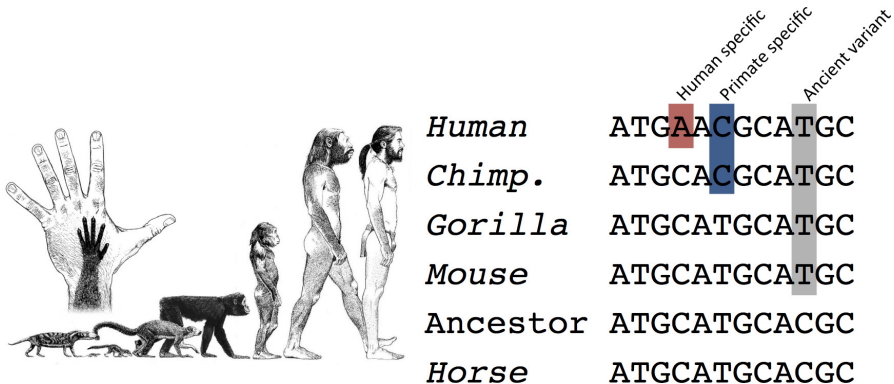
b YCHATTGTTCTC

c

A	002700000010
C	464100000505
G	000001800112
T	422087088261



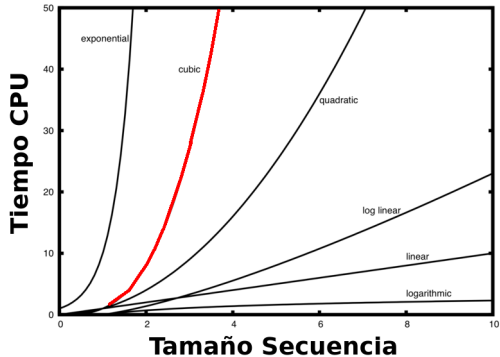
Evolutionary Approach



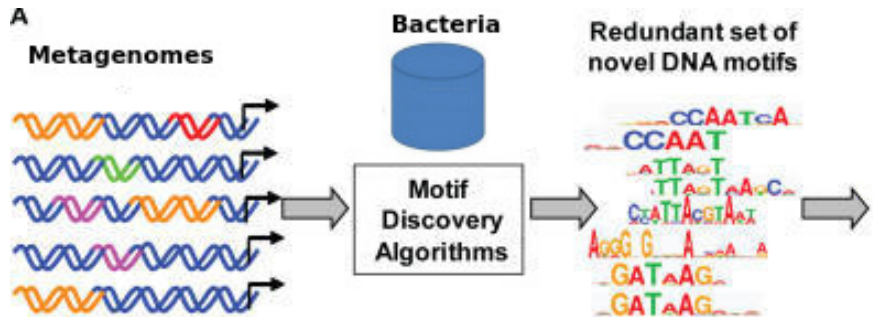
Resultados

Algoritmos no Funcionan Bien con Secuencias Largas

- Word-based Method: **Weeder**
- Probabilistic Method: **Meme**

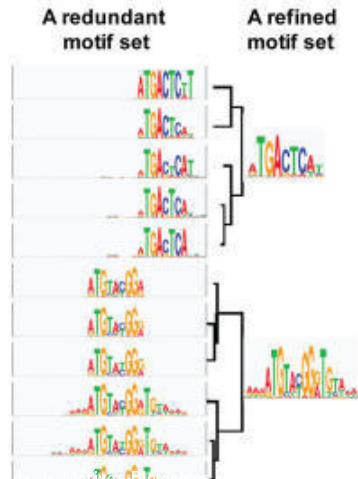


Enfoque Evolutivo - Parte A

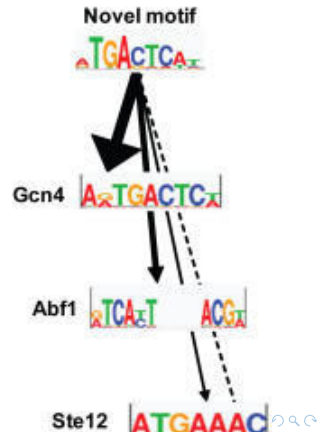


Enfoque Evolutivo - Parte B

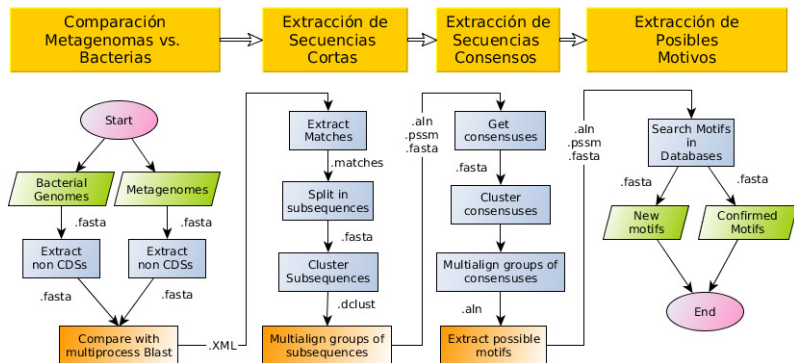
Ⓑ Cluster & merge redundant motifs



Ⓒ Relating novel motifs to known TF binding sites



Pipeline Para Búsqueda de Motivos en Regiones No-Codificantes



Bases de Datos para Validar los Motivos

- **Prokaryotes (Prodoric Release 8.9):** Prokaryotic motifs from PRODORIC downloaded from <http://prodoric.tu-bs.de/msearch.php>.
- **Prokaryotes (RegTransBase v4):** Manually curated prokaryotic motifs from RegTransBase; downloaded from <http://regtransbase.lbl.gov/cgi-bin/regtransbase?page=downloads>.
- **E. coli ******(DPINTERACT):** Footprinting databases based on DPINTERACT; downloaded from http://arep.med.harvard.edu/ecoli_matrices.

Resultados Preliminares

prodoric.dbout											
1	#pattern	name	sequence name			start	stop	strand	score	p-value	q-value
matched sequence											
2	MX000001	3050	2	16	+	6.36429	5.88e-05		GGTGATGCTGCCAAC		
3	MX000002	2115	9	22	+	9.7156	4.57e-05		CTGATACACATCAG		
4	MX000002	2115	9	22	-	11.8257	2.31e-05		CTGATGTGTATCAG		
5	MX000003	2740	3	9	-	9.10714	6.51e-05		TACCCAT		
6	MX000003	2987	12	18	-	9.10714	6.51e-05		TACCCAT		
7	MX000003	3191	15	21	+	9.10714	6.51e-05		TACCCAT		
8	MX000003	6651	1	7	-	9.10714	6.51e-05		TACCCAT		
9	MX000003	7515	1	7	-	9.10714	6.51e-05		TACCCAT		
10	MX000003	8395	2	8	-	9.10714	6.51e-05		TACCCAT		
11	MX000003	11611	6	12	-	9.10714	6.51e-05		TACCCAT		
12	MX000003	12615	3	9	+	9.10714	6.51e-05		TACCCAT		
13	MX000004	2780	11	24	+	11.0444	6.22e-05		TTGAAGTACCTCAA		
14	MX000005	2705	8	23	+	7.90826	1.82e-05		TGGGGAATATTGCACA		
15	MX000005	5487	1	16	+	9.58716	1.06e-05		TGTGCGGTATTTCACA		
16	MX000005	5487	1	16	-	17.1927	4.31e-07		TGTGAAATACCGCACA		
17	MX000006	0562	8	17	+	10.73	2.92e-05		GGGACAAGCC		
18	MX000006	0735	12	21	-	10.73	2.92e-05		GGGACAAGCC		
19	MX000006	1058	8	17	-	10.73	2.92e-05		GGGACAAGCC		
20	MX000006	1078	7	16	+	10.73	2.92e-05		GGGACAAGCC		
21	MX000006	2292	5	14	+	12.15	1.65e-06		TGAACAAGCG		
22	MX000006	4543	10	19	+	10.28	6.07e-05		GAAACAAGCG		
23	MX000006	5237	1	10	-	10.56	3.73e-05		GAGACAAGCG		
24	MX000006	5466	2	11	+	10.46	4e-05	TGAACAATCG			
25	MX000006	6164	1	10	-	10.02	8.58e-05		GAGACAAGCA		
26	MX000006	7517	2	11	+	10.2	6.42e-05		TGGACAATCA		
27	MX000006	13229	2	11	+	10.16	7.36e-05		TGGACAAACC		
28	MX000006	13463	2	11	+	9.99	9.44e-05		AGACCAAGCG		
29	MX000006	13912	1	10	-	10.16	7.36e-05		TGGACAAACC		
30	MX000007	0305	5	21	+	11.2626	5.29e-05		CGTCCCAAGTGGTAAG		

25,1

Top

25,1

Top

Gracias!