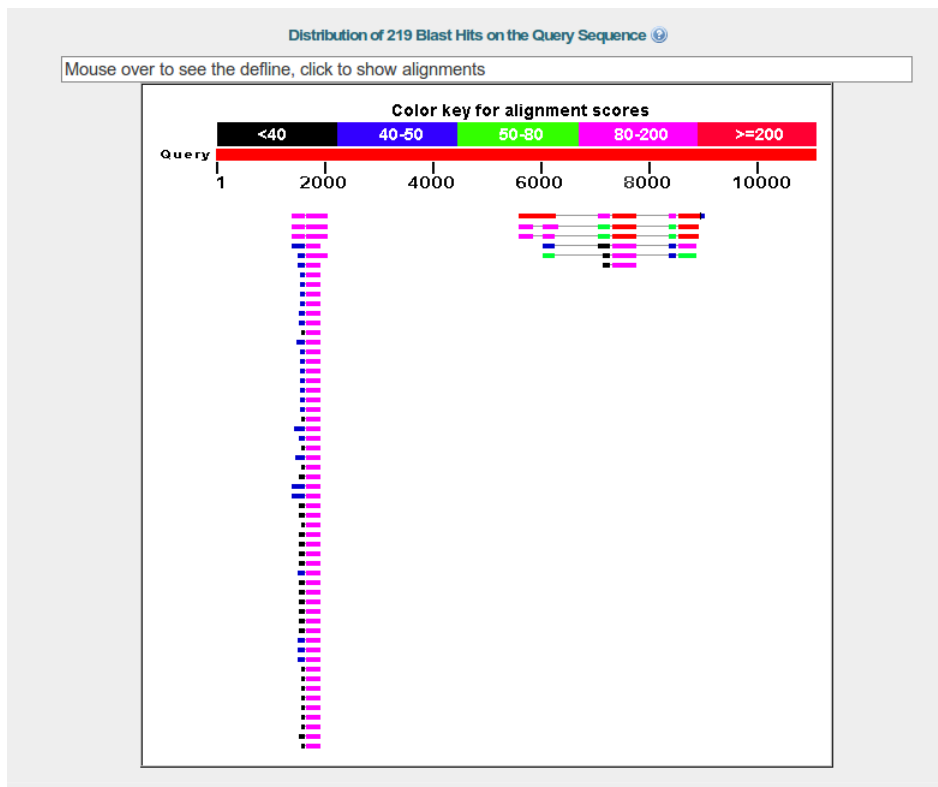


Como un primer acercamiento para la identificación de genes en la secuencia genómica podemos realizar un blastx y comparar contra las proteínas conocidas. Vamos a hacer un Blastx contra la base de proteínas Refseq de la NCBI

<http://blast.ncbi.nlm.nih.gov/>. Programa blastx, base de datos refseq.

El resultado gráfico es el siguiente:



Se pueden identificar dos regiones que dan positivo:

- 1- En las posiciones cercanas a 2000 con 2 HSP cada uno
- 2- Entre 5000 y 9000 varios hit con 5-6 HSP cada uno.

Veamos la tabla

Sequences producing significant alignments:							
Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
NP_510440.1	defective HatCHing family member (hch-1) [Caenorhabditis elegans]	301	1065	17%	1e-78	100%	<a href="#">U</a> <a href="#">G</a>
NP_510439.1	SOX (mammalian SRY box) family member (sox-3) [Caenorhabditis elegans]	187	347	5%	1e-85	96%	<a href="#">U</a> <a href="#">G</a>
XP_003106263.1	CRE-SOX-3 protein [Caenorhabditis remanei]	177	321	5%	7e-78	84%	<a href="#">G</a>
XP_002645839.1	C. briggsae CBR-SOX-3 protein [Caenorhabditis briggsae]	172	312	5%	2e-75	83%	<a href="#">G</a> <a href="#">G</a>
XP_003106283.1	CRE-HCH-1 protein [Caenorhabditis remanei]	224	856	15%	2e-72	85%	<a href="#">G</a> <a href="#">G</a>
XP_002645840.1	C. briggsae CBR-HCH-1 protein [Caenorhabditis briggsae]	224	849	15%	3e-72	88%	<a href="#">G</a> <a href="#">G</a>
XP_003098250.1	hypothetical protein CRE_08443 [Caenorhabditis remanei]	135	345	11%	1e-33	49%	<a href="#">G</a> <a href="#">G</a>
XP_001893910.1	HMG box family protein [Brugia malayi]	118	164	4%	5e-31	73%	<a href="#">G</a> <a href="#">G</a>
XP_003148818.1	HMG box family protein [Loa loa]	119	163	4%	1e-30	46%	<a href="#">G</a> <a href="#">G</a>
NP_999638.1	transcription factor SoxB2 [Strongylocentrotus purpuratus]	117	162	3%	1e-30	71%	<a href="#">U</a> <a href="#">G</a> <a href="#">M</a>
XP_001972711.1	GG13735 [Drosophila erecta]	120	160	3%	5e-30	71%	<a href="#">G</a>
XP_002094767.1	GE20029 [Drosophila yakuba]	120	160	3%	5e-30	71%	<a href="#">G</a>
NP_648694.1	Sox21a [Drosophila melanogaster]	120	160	3%	5e-30	71%	<a href="#">U</a> <a href="#">G</a>
XP_002030464.1	GM24557 [Drosophila sechellia]	120	160	3%	5e-30	71%	<a href="#">U</a> <a href="#">G</a>
XP_002593017.1	hypothetical protein BRAFLDRAFT_263176 [Branchiostoma floridae]	117	160	3%	6e-30	71%	<a href="#">G</a>
XP_971910.1	PREDICTED: similar to Sox21a CG7345-PA [Tribolium castaneum]	118	160	3%	7e-30	67%	<a href="#">U</a> <a href="#">G</a> <a href="#">M</a>
XP_002084834.1	GD12630 [Drosophila simulans]	120	160	2%	9e-30	75%	<a href="#">U</a> <a href="#">G</a>
NP_001158461.1	Sox14/21-like protein [Saccoglossus kowalevskii]	117	159	3%	1e-29	72%	<a href="#">U</a> <a href="#">G</a> <a href="#">M</a>
XP_002021381.1	GL24833 [Drosophila persimilis]	118	159	3%	2e-29	70%	<a href="#">G</a>
XP_002061714.1	GK17145 [Drosophila willistoni]	118	159	3%	2e-29	70%	<a href="#">G</a>
XP_001353546.2	GA20281 [Drosophila pseudoobscura pseudoobscura]	118	159	3%	2e-29	70%	<a href="#">G</a> <a href="#">G</a>
XP_002008931.1	GI11538 [Drosophila mojavensis]	118	159	2%	2e-29	71%	<a href="#">G</a> <a href="#">G</a>
XP_001984256.1	GH16347 [Drosophila grimshawi]	118	159	2%	2e-29	71%	<a href="#">G</a> <a href="#">G</a> <a href="#">G</a>
XP_002048055.1	GJ11557 [Drosophila virilis]	118	159	2%	2e-29	71%	<a href="#">G</a> <a href="#">G</a>
XP_001956434.1	GF25206 [Drosophila ananassae]	117	158	2%	3e-29	71%	<a href="#">G</a> <a href="#">G</a>
XP_002424557.1	predicted protein [Pediculus humanus corporis]	119	157	2%	5e-29	73%	<a href="#">G</a>

La mayoría de hits dan valores de e-value muy bajos y significativos. Podemos observar los mejores hits que son :

- 1- defective HatCHing family member (hch-1) [Caenorhabditis elegans]
- 2- SOX (mammalian SRY box) family member (sox-3) [Caenorhabditis]

Veamos sus alineamientos

>[ref|NP\\_510440.1|](#) **UG** defective HatCHing family member (hch-1) [Caenorhabditis elegans]  
Length=605

[GENE ID: 181564 hch-1](#) | defective HatCHing [Caenorhabditis elegans]  
(10 or fewer PubMed links)

this subject sequence by: Sort alignments for  
E value Score  
Percent identity Query start position

Subject start position

Score = 268 bits (685), Expect(2) = 1e-90  
Identities = 132/147 (90%), Positives = 132/147 (90%), Gaps = 15/147 (10%)  
Frame = -1

Query	7788	CYSNIGKVS RFPQDVSIGWGCTSLGTVCHEIGKRPKPK*IDTTHFSGHALGFYHEQARYD	7609	
		CYSNIGKVS RFPQDVSIGWGCTSLGTVCHEIG	HALGFYHEQARYD	
Sbjct	191	CYSNIGKVS RFPQDVSIGWGCTSLGTVCHEIG-----	HALGFYHEQARYD	235
Query	7608	RDDYVSILTQNIQDMYLSQFTKQSASSMVDYGVGYDYGSVMHYDQAAFSSTGGNTIATRD	7429	
		RDDYVSILTQNIQDMYLSQFTKQSASSMVDYGVGYDYGSVMHYDQAAFSSTGGNTIATRD		
Sbjct	236	RDDYVSILTQNIQDMYLSQFTKQSASSMVDYGVGYDYGSVMHYDQAAFSSTGGNTIATRD	295	
Query	7428	PNFQATIGQRVAPSFADV KRINFAYCN	7348	
		PNFQATIGQRVAPSFADV KRINFAYCN		
Sbjct	296	PNFQATIGQRVAPSFADV KRINFAYCN	322	

Score = 95.5 bits (236), Expect(2) = 1e-90  
Identities = 70/72 (98%), Positives = 72/72 (100%), Gaps = 0/72 (0%)  
Frame = -3

Query	7303	SATCSNYLDCQNGGYInpndcnnckcpPGFGGQLCDVAGTNSNGCGAGDitatssiqtis	7124
		++TCSNYLDCQNGGYINPNDCNNCKCPPGFGGQLCDVAGTNSNGCGAGDITATSSIQTIS	
Sbjct	322	NSTCSNYLDCQNGGYINPNDCNNCKCPPGFGGQLCDVAGTNSNGCGAGDITATSSIQTIS	381
Query	7123	asgALTCNYVIK	7088
		ASGALTCNYVIK	
Sbjct	382	ASGALTCNYVIK	393

Score = 263 bits (671), Expect(2) = 4e-87  
Identities = 129/132 (98%), Positives = 129/132 (98%), Gaps = 0/132 (0%)  
Frame = -1

Query	8961	ILFQKSYFADFVNGKGPFKQADALKFMDKMTILNKLQADILGIPQPDEFSALDFEDKIES	8782
		I KSYFADFVNGKGPFKQADALKFMDKMTILNKLQADILGIPQPDEFSALDFEDKIES	
Sbjct	16	ICHAKSYFADFVNGKGPFKQADALKFMDKMTILNKLQADILGIPQPDEFSALDFEDKIES	75

```

Query 8781 KPDEIPYLFEGDMVLTDQMDLIIKNVRDQYWARKSSTNEFLYAIRGKRSMSTFLSERWS 8602
Sbjct 76 KPDEIPYLFEGDMVLTDQMDLIIKNVRDQYWARKSSTNEFLYAIRGKRSMSTFLSERWS 135

Query 8601 FPVPPYYIDTSSG 8566
Sbjct 136 FPVPPYYIDTSSG 147

```

Score = 89.0 bits (219), Expect(2) = 4e-87  
Identities = 43/43 (100%), Positives = 43/43 (100%), Gaps = 0/43 (0%)  
Frame = -3

```

Query 8521 VNTNAVLAVGAKWEQETCARFTRLNSYSSSSRQNALRFISGNG 8393
Sbjct 148 VNTNAVLAVGAKWEQETCARFTRLNSYSSSSRQNALRFISGNG 190

```

Score = 301 bits (771), Expect = 1e-78  
Identities = 212/232 (92%), Positives = 215/232 (93%), Gaps = 15/232 (6%)  
Frame = -2

```

Query 6305 HHFFQAPVGAQVYFQMTAATFSRYSPCTTNYLEINYGRDFSRVGRFCASYPTISLSETN 6126
Sbjct 389 ++ +APVGAQVYFQMTAATFSRYSPCTTNYLEINYGRDFSRVGRFCASYPTISLSETN 448

```

```

Query 6125 TLVVIYKGVNGARFSLNYRYDpvtfstsaptttstttttapitvptvspttttrqtttt 5946
Sbjct 449 TLVVIYKGVNGARFSLNYRYDPVTFSTAPTTTTSTTTTTAPITVPTVSPTTTTTRQTTTT 508

```

```

Query 5945 artsttttttqappttttstSQCASWSACSAQCGGCGTQSRR*LNAIYSDTYFMNYRCGT 5766
Sbjct 509 ARTSTTTTTTQAPTTTTSTSQCASWSACSAQCGGCGTQSRR CGT 553

```

```

Query 5765 YVETVYCNTNPCTGGYCCRPFFYVTSFGTGYCRRPGADTPAAPQRYVEQRKG 5610
Sbjct 554 YVETVYCNTNPCTGGYCCRPFFYVTSFGTGYCRRPGADTPAAPQRYVEQRKG 605

```

Score = 48.1 bits (113), Expect = 0.023  
Identities = 22/30 (74%), Positives = 24/30 (80%), Gaps = 0/30 (0%)  
Frame = -2

```

Query 9053 MVSYPVLLIVLCLLPICHA VSCVSVFLNKK 8964
Sbjct 1 MVSYPVLLIVLCLLPICHA S + +N K 30

```

Presenta 6 HSP todos ellos con alta similitud, pero si nos fijamos algunos presentan indels internos y algunos extremos de los alineamientos no acaban con tanta similitud.

Podemos identificar varios exones correspondientes a estos HSP:

1 HSP: Nos indica que la pauta traducida es la -1, con lo que el gen esta situado en la cadena complementaria. Contiene un indel de unos 15-16 aminoácidos que no están presentes en la proteína. Lo que sucede es que hay dos exones separados por un intrón pequeño; pero que la pauta en el genómico es la misma (ambos la -1) y el Blast los ha unido porque mejoraba el alineamiento.

De modo: 2 exones 7788-7692 y de 7647-7348 (posiciones aproximadas he contado a ojo los nt)

2 HSP: En este caso corresponde a un solo exón, pero se ve que el extremo no acaba bien.  
Exón: 7303-7088

3 HSP En este caso corresponde a un solo exón, pero se ve que el extremo no acaba bien.  
Exón: 8961-8566

4 HSP: exón 8521-8393

5 HSP: Como rn rl HSP 1: 2 exones en la pauta -2. exon 6305-5819 y 5760-5610

6 HSP: 1 exón con cortes no claros 9053-8964.

En este caso los cortes se ven que están mal y que es que el programa ha extendido mas los HSP por los intrones, gracias a que la similitud es muy grande ya el genómico y la proteína son de la misma especie. En otros casos es mas complicado y no sabemos el punto de corte exacto (ver hits con otras especies más alejadas).

Es un gen con 8 exones y codifica para HatCHing family member. Los puntos de corte de los exones son erroneos y no tenemos informacion sobre los 3' y 5' UTR.

Que sucede con el segundo gen:

>ref|NP\_510439.1| **UG**SOX (mammalian SRY box) family member (sox-3)  
[Caenorhabditis  
elegans]  
Length=212

[GENE ID: 185534 sox-3](#) | SOX (mammalian SRY box) family [Caenorhabditis elegans]  
(10 or fewer PubMed links)

this subject sequence by: Sort alignments for  
E value Score  
Percent identity Query start position  
Subject start position

Score = 187 bits (474), Expect(2) = 1e-85  
Identities = 110/129 (86%), Positives = 110/129 (86%), Gaps = 19/129 (14%)  
Frame = -2

Query	2084	MTDLSCLYPSLLCTEAAKTSYDEDTTsvssglsppgspVDLQNSLDHVKRPMAFMVWSR	1905
		MTDLSCLYPSLLCTEAAKTSYDEDTTSVSSGLSPPGSPVDLQNSLDHVKRPMAFMVWSR	
Sbjct	1	MTDLSCLYPSLLCTEAAKTSYDEDTTSVSSGLSPPGSPVDLQNSLDHVKRPMAFMVWSR	60
Query	1904	GQRRKMAQVHFYLNFLGYVSK*FIQDNPKMHNSEISKRLGAEWKQLSEQEKRPFIDEA	1725
		GQRRKMA QDNPKMHNSEISKRLGAEWKQLSEQEKRPFIDEA	
Sbjct	61	GQRRKMA-----QDNPKMHNSEISKRLGAEWKQLSEQEKRPFIDEA	101
Query	1724	KRLRALHMK 1698	
		KRLRALHMK	
Sbjct	102	KRLRALHMK 110	

Score = 160 bits (404), Expect(2) = 1e-85  
Identities = 75/78 (97%), Positives = 77/78 (99%), Gaps = 0/78 (0%)  
Frame = -1

```

Query 1656 YTQEHPDYKYRPRRKPKSSNLKQQPRLNIAMPTIPPQSLFNYSTAFDSLKTHDLSQYYSS 1477
+ +EHPDYKYRPRRKPKSSNLKQQPRLNIAMPTIPPQSLFNYSTAFDSLKTHDLSQYYSS
Sbjct 108 HMKEHPDYKYRPRRKPKSSNLKQQPRLNIAMPTIPPQSLFNYSTAFDSLKTHDLSQYYSS 167

Query 1476 FFQSPVLSGSTYAPYNMM 1423
FFQSPVLSGSTYAPYNMM
Sbjct 168 FFQSPVLSGSTYAPYNMM 185

```

Siguiendo el mismo razonamiento: 3 exones en la cadena complementaria.

2084-1883, 1826-1698., 1656-1423

Y codificaría para una SOX (mammalian SRY box) family member (sox-3) [Caenorhabditis elegans]

Bueno ya hemos identificado dos genes en la región, pero su estructura (exones, intrones, UTR, etc) es imprecisa. Podemos mejorar comparando los cDNAs y ESTs con el genómico. Con este alineamiento si que obtendríamos las posiciones exactas de los exones y de las regiones no codificantes, obviamente únicamente de aquellas presentes en los cDNAs. Para ello podemos utilizar el Est2genome.

Sin resultado cDNA1, cDNA2, cDNA4, est2, est4 y est6

Con resultado:

### CDNA3

Feature	Start	End	Score	Start	End	Score	Start	End	Score
Exon	276	100.0	5500	5775	5775	1	276	100.0	5500
-Intron	-20	0.0	5776	5820	5820				
Exon	470	100.0	5821	6290	6290	277	746	100.0	5821
-Intron	-20	0.0	6291	7087	7087				
Exon	212	100.0	7088	7299	7299	747	958	100.0	7088
-Intron	-20	0.0	7300	7346	7346				
Exon	303	100.0	7347	7649	7649	959	1261	100.0	7347
-Intron	-20	0.0	7650	7694	7694				
Exon	95	100.0	7695	7789	7789	1262	1356	100.0	7695
-Intron	-20	0.0	7790	8393	8393				
Exon	127	100.0	8394	8520	8520	1357	1483	100.0	8394
-Intron	-20	0.0	8521	8564	8564				
Exon	385	100.0	8565	8949	8949	1484	1868	100.0	8565
-Intron	-20	0.0	8950	8996	8996				
Exon	69	100.0	8997	9065	9065	1869	1937	100.0	8997

8 exones en la posición que habíamos detectado la HatCHing mediante blast. Comparar y comprobar la coincidencia

Blast (puesto en cadena complementaria):

5610-5760  
5819-6305  
7088-7303  
7348- 7647  
7692-7788  
8393- 8521  
8566-8961

9053-8964.

A parte de separar bien los intrones y exones, en el EST2genome se extienden más ya que el cDNA contiene las regiones 5' y 3' UTR no presentes en las proteínas.

### EST1

Note Best alignment is between forward est and forward genome, but splice sites imply REVERSED GENE

Exon	281	98.6	1359	1647	genomico_problema	1	289	EST1
-Intron	-20	0.0	1648	1697	genomico_problema			
Exon	122	98.4	1698	1823	genomico_problema	290	415	EST1
-Intron	-20	0.0	1824	1880	genomico_problema			
Exon	28	100.0	1881	1908	genomico_problema	416	443	EST1
Span	391	98.6	1359	1908	genomico_problema	1	443	EST1
Segment	281	98.6	1359	1647	genomico_problema	1	289	EST1
Segment	122	98.4	1698	1823	genomico_problema	290	415	EST1
Segment	28	100.0	1881	1908	genomico_problema	416	443	EST1

### EST3

Note Best alignment is between reversed est and forward genome, but splice sites imply REVERSED GENE

Exon	237	100.0	1411	1647	genomico_problema	1	237	est3
-Intron	-20	0.0	1648	1697	genomico_problema			
Exon	126	100.0	1698	1823	genomico_problema	238	363	est3
-Intron	-20	0.0	1824	1880	genomico_problema			
Exon	176	100.0	1881	2056	genomico_problema	364	539	est3
Span	499	100.0	1411	2056	genomico_problema	1	539	est3
Segment	237	100.0	1411	1647	genomico_problema	1	237	est3
Segment	126	100.0	1698	1823	genomico_problema	238	363	est3
Segment	176	100.0	1881	2056	genomico_problema	364	539	est3

Los EST3 y 5 se corresponden con la region del gen SOX detectado por Blast. La estructura del gen según estos est seria:

1359-1647; 1698-1823; 1881 2056

Pero en este caso esta formada por el alineamiento de 2 EST y no sabemos si esta completo. Al comparar con el alineamiento del Blastx vemos que coincide las regiones y que según este el ultimo exon se expande hasta la posición 2084 en vez de 2056, eso es posiblemente porque los EST no cubren esa región.

1423-1656

1698-1826

1883-2084

### EST5

Align EST sequences to genomic DNA sequence

Note Best alignment is between forward est and forward genome, and splice sites imply forward gene

Exon 222 100.0 10879 11100 genomico\_problema 4 225 est5  
 Span 222 100.0 10879 11100 genomico\_problema 4 225 est5  
 Segment 222 100.0 10879 11100 genomico\_problema 4 225 est5

Este EST5 se alinea en una región no identificada con el Blastx. El EST5 alinea justo en el extremo del genómico y no cubre toda su extensión. Podemos hacer un Blastx con el EST5 para ver si podemos identificar a que se parece.

### La tabla de resultados

Sequences producing significant alignments:							
Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
<a href="#">NP_510441.1</a>	hypothetical protein F40E10.5 [Caenorhabditis elegans]	<a href="#">155</a>	356	81%	1e-83	100%	<a href="#">U</a> <a href="#">G</a>
<a href="#">XP_003106250.1</a>	hypothetical protein CRE_15305 [Caenorhabditis remanei]	<a href="#">145</a>	331	89%	3e-76	91%	<a href="#">G</a>
<a href="#">XP_002645841.1</a>	Hypothetical protein CBG07574 [Caenorhabditis briggsae]	<a href="#">142</a>	318	81%	2e-72	91%	<a href="#">G</a>
<a href="#">XP_002634376.1</a>	C. briggsae CBR-ORA-1 protein [Caenorhabditis briggsae]	<a href="#">57.0</a>	88.2	53%	6e-10	42%	<a href="#">G</a>
<a href="#">XP_003108375.1</a>	CRE-ORA-1 protein [Caenorhabditis remanei]	<a href="#">56.6</a>	87.8	53%	7e-10	42%	<a href="#">G</a>
<a href="#">XP_003147615.1</a>	hypothetical protein LOAG_12053 [Loa loa]	<a href="#">60.8</a>	60.8	35%	9e-08	38%	<a href="#">G</a>

El primer hit:

```
>ref|NP_510441.1| UniGene info linked to NP_510441.1Gene info linked to
NP_510441.1 hypothetical protein F40E10.5 [Caenorhabditis elegans]
Length=284
```

GENE ID: 185535 F40E10.5 | hypothetical protein [Caenorhabditis elegans]

this subject sequence by: Sort alignments for  
E value Score  
 Percent identity Query start position  
 Subject start position  
 Score = 155 bits (391), Expect(3) = 1e-83  
 Identities = 73/74 (99%), Positives = 74/74 (100%), Gaps = 0/74 (0%)  
 Frame = +2

```
Query 224 KTKPCGIPKFINELPSEAAEKITKIWEKYVEGNTCEKEHEETRNIIRGLTTEERDKVFAG 403
+TKPCGIPKFINELPSEAAEKITKIWEKYVEGNTCEKEHEETRNIIRGLTTEERDKVFAG
Sbjct 34 RTKPCGIPKFINELPSEAAEKITKIWEKYVEGNTCEKEHEETRNIIRGLTTEERDKVFAG 93
```

```
Query 404 RCGPSFLKNVSSTV 445
RCGPSFLKNVSSTV
Sbjct 94 RCGPSFLKNVSSTV 107
```

Score = 122 bits (305), Expect(3) = 1e-83  
 Identities = 61/61 (100%), Positives = 61/61 (100%), Gaps = 0/61 (0%)  
 Frame = +3

```
Query 447 REEFKNVWFDYKLSVVEKELALQKLAYSLLTGESLALFNKWEELKTRKVEVARRLSELS 626
REEFKNVWFDYKLSVVEKELALQKLAYSLLTGESLALFNKWEELKTRKVEVARRLSELS
Sbjct 108 REEFKNVWFDYKLSVVEKELALQKLAYSLLTGESLALFNKWEELKTRKVEVARRLSELS 167
```

```
Query 627 P 629
P
Sbjct 168 P 168
```

Score = 79.3 bits (194), Expect(3) = 1e-83

Identities = 34/36 (95%), Positives = 35/36 (98%), Gaps = 0/36 (0%)  
 Frame = +1

```
Query 76 MQTLILAIIFLPCIVLAFPPRHSDPFNPLGPDWHKRV 183
          MQTLILAIIFLPCIVLAFPPRHSDPFNPLGPDWHKR +
Sbjct 1  MQTLILAIIFLPCIVLAFPPRHSDPFNPLGPDWHKRTK 36
```

Se parece a una proteína hipotética de *C. elegans*. Con el genómico no la detectamos porque la región cubierta en el genómico es pequeña.

Con la comparación de los cDNAs hemos detectado genes nuevos y delimitado mejor la estructura.

Podemos utilizar también el genemark, ya hemos visto que el organismo es *C. elegans*. Así que usando GeneMark. El fichero gráfico lo teneis colgado en la Red.

```
Eukariotyc GeneMark.hmm version bp 3.9e Dec 2010
Sequence name: Tue Feb 22 10:46:42 EST 2011
Sequence length: 11100 bp
G+C content: 36.00%
Matrices file: /home/genemark/euk_ghm.matrices/celegans_hmm3.0mod
Tue Feb 22 10:46:42 2011
```

#### Predicted genes/exons

Gene #	Exon #	Strand	Exon Type	Exon Range	Exon Length	Start/End Frame
1	2	-	Terminal	62 93	32	3 2 - -
	1	-	Initial	1229 1256	28	1 1 - -
2	4	-	Terminal	1339 1647	309	3 1 - -
	3	-	Internal	1698 1823	126	3 1 - -
	2	-	Internal	1881 2088	208	3 3 - -
	1	-	Initial	4224 4243	20	2 1 - -
3	8	-	Terminal	5607 5775	169	3 3 - -
	7	-	Internal	5821 6290	470	2 1 - -
	6	-	Internal	7088 7299	212	3 2 - -
	5	-	Internal	7347 7649	303	1 2 - -
	4	-	Internal	7695 7789	95	1 3 - -
	3	-	Internal	8394 8520	127	2 2 - -
	2	-	Internal	8565 8949	385	1 1 - -
1	-	Initial	8997 9053	57	3 1 - -	

El programa detecta 3 genes, el 2 y 3 corresponden a los genes detectados mediante Blast. El 1 es nuevo. Está en una región que no ha dado similitud ni en la base de datos, ni con los cDNAs y ,además, son exones muy pequeños y las probabilidades no son muy altos. Lo mas probable es que no sea cierto, pero habría que tener en cuenta esa posibilidad.

Ahora habría que comparar los resultados de los tres sistemas y desarrollar un modelo de la región. Ya que estos tres sistemas son complementarios y entre ellos se suplen las deficiencias.