

UTILIZZO DI BLAST PER ALCUNE SEMPLICI APPLICAZIONI IN STUDI GENOMICI

- Come prima cosa diamo un'occhiata alla nostra sequenza di interesse, chiamata «unknown sequence»
- Con un doppio click possiamo visualizzarla
- Si tratta di una sequenza nucleotidica lunga poco più di 800 nucleotidi, corrispondente ad un mRNA di ostrica la cui funzione è ignota

The screenshot displays the CLC Genomics Workbench 11.0 interface. The main window shows a DNA sequence of approximately 800 nucleotides, with several restriction enzyme sites highlighted: HindIII, EcoRI, XhoI, and PstI. The sequence is displayed in a multi-line format with nucleotide positions indicated above each line. The left sidebar shows a project tree with 'unknown_sequence' selected. The bottom left shows a 'Toolbox' with 'SRA Download' and 'Searching SRA Database...' options. The right sidebar shows 'Sequence Settings' with various options like 'Spacing layout', 'No spacing', 'Auto wrap', and 'Restriction sites'.

STEP 1: ANNOTAZIONE FUNZIONALE

- Per capire quale può essere la funzione del mRNA ignoto potrebbe essere utile capire prima di tutto quale sia la proteina da esso codificata
- Questo è possibile grazie ad il confronto con sequenze la cui funzione è nota che sono state precedentemente depositate in un database pubblicamente disponibile sul server dell'NCBI, un grande centro bioinformatico americano
- Possiamo collegarci all'indirizzo <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Come prima cosa dobbiamo selezionare il tipo di BLAST più appropriato, nel nostro caso BLASTx (la sequenza nucleotidica deve essere tradotta per essere confrontata con un database proteico)

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

IgBLAST 1.8.0 released

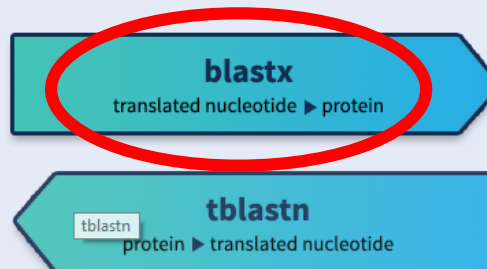
A new version of IgBLAST is now available.
Wed, 15 Nov 2017 16:00:00 EST

[More BLAST news...](#)

Web BLAST

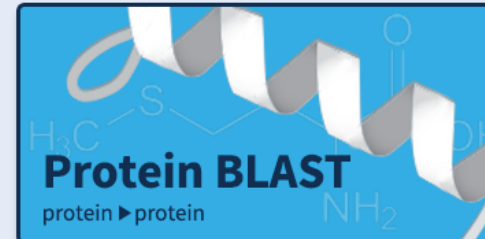


Nucleotide BLAST
nucleotide ▶ nucleotide



blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide



Protein BLAST
protein ▶ protein

BLAST Genomes

Search

Human

Mouse

Rat

Microbes

Translated BLAST: blastx

blastn blastp **blastx** tblastn tblastx

BLASTX search protein databases using a translated nucleotide query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

GTTATCACCCGCATTCCGGAACATTTATCGCAACTCGGTCTGGATATTATGCTTTACGTGGTCCTTTAGGAT
TCAAGGTGATGCTCATATTTGACCGCAACTCGTGGTAAATGACACGCCTAGAGGTACCGTTTACTATGACGCA
ACAGAGAAAGTAGGTGGAAACACTGGTGGGACGGTTGTGCTTCACGTTAACCAGGATGATGAAGTCTTTATAC
GTACAACCAATACTAACTATAACCTTGGTGACATTTTGAGTGATAACGCAGSAAGAAGCTATTTTGGTGGTTG
GCTTTTATCCTAA

Clear Query subrange

From
To

INCOLLIAMO LA SEQUENZA NUCLEOTIDICA NEL RIQUADRO

Or, upload file

Scegli file Nessun file selezionato

Genetic code

Standard (1)

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

- UniProtKB/Swiss-Prot(**swissprot**)
- Non-redundant protein sequences (nr)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot(**swissprot**)
- Patented protein sequences(pat)
- Protein Data Bank proteins(pdb)
- Metagenomic proteins(env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

Organism
Optional

Exclude
Optional

Entrez Query
Optional

ASSICURIAMOCI DI SELEZIONARE DAL MENU' A TENDINA IL DATABASE «UNIPROTKB» CHE CONTIENE SEQUENZE PROTEICHE LA CUI FUNZIONE E' STATA DETERMINATA SPERIMENTALMENTE E DI CUI CI POSSIAMO FIDARE

BLAST

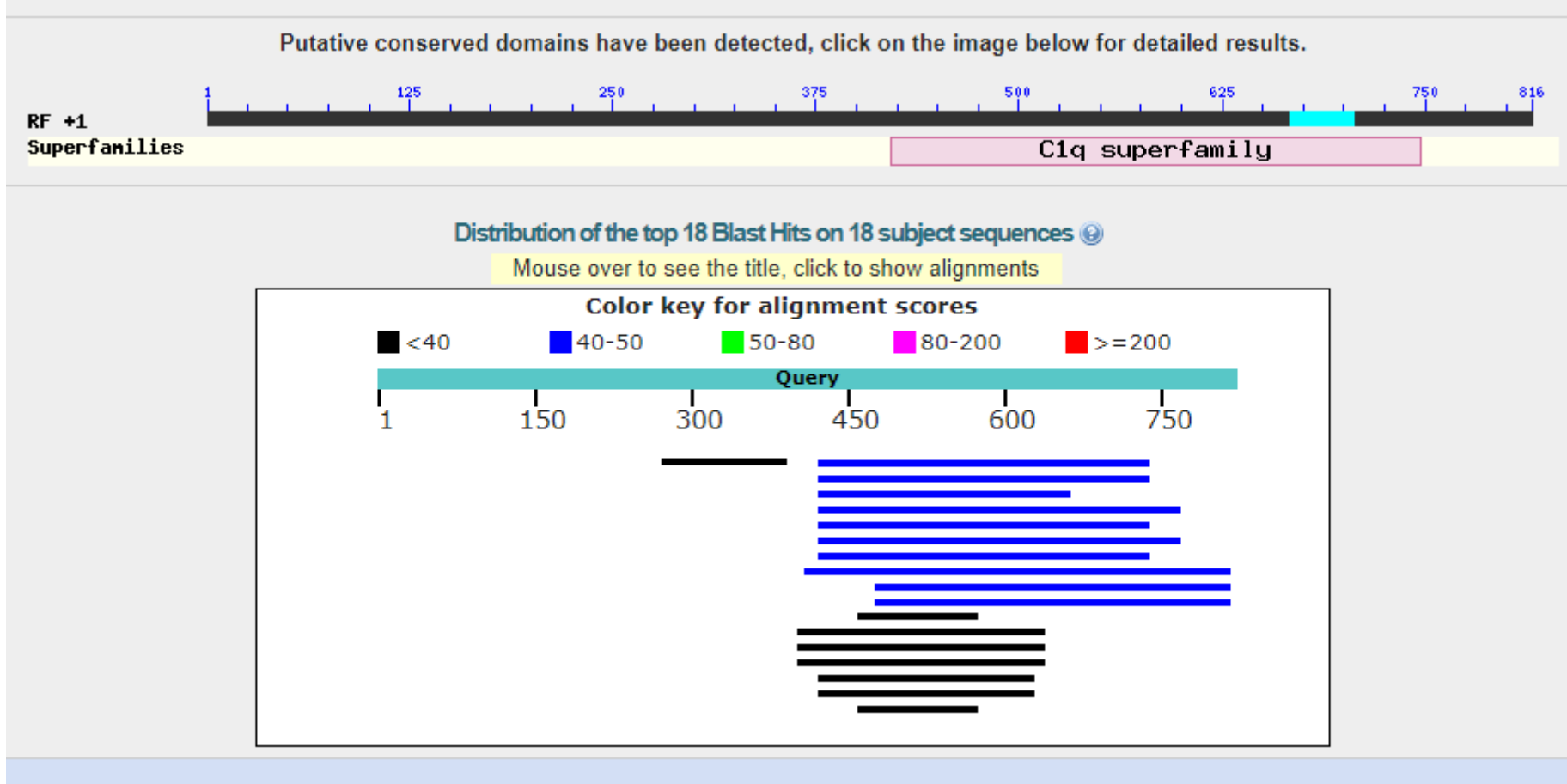
Search database UniProtKB/Swiss-Prot(**swissprot**) using Blastx (search protein databases using a translated nucleotide query)

Show results in a new window

CLICCHIAMO SUL TASTO «BLAST» PER LANCIARE LA RICERCA

[Algorithm parameters](#)

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign



Nel giro di pochi secondi dovvemmo ottenere i risultati, organizzati in tre sezioni

- 1) Una parte grafica con barre colorate
- 2) Una tabella con la lista dei risultati
- 3) Il dettaglio degli allineamenti ottenuti tra la sequenza query (input) e subject (quelle trovate nel database)

I risultati sono ordinati per e-value dal più significativo al meno significativo

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

		Alignments	Download	GenPept	Graphics						
	Description	Max score	Total score	Query cover	E value	Ident	Accession				
<input type="checkbox"/>	RecName: Full=Complement C1q-like protein 3; AltName: Full=C1q and tumor necrosis factor-related protein 13; Short=C1q/TNF-related protein 13; Short=CTRP13; AltName: Full=Gliacolin; Flags: Precursor	43.1	43.1	38%	0.001	31%	Q9ESN4.1				
<input type="checkbox"/>	RecName: Full=Complement C1q-like protein 3; AltName: Full=C1q and tumor necrosis factor-related protein 13; Short=C1q/TNF-related protein 13; Flags: Precursor	43.1	43.1	38%	0.001	31%	Q5VWW1.1				
<input type="checkbox"/>	RecName: Full=Heavy metal-binding protein HIP	42.4	42.4	29%	0.001	29%	P83425.1				
<input type="checkbox"/>	RecName: Full=Complement C1q-like protein 4; AltName: Full=C1q and tumor necrosis factor-related protein 11; Short=C1q/TNF-related protein 11; Short=C1qTNF11; Short=CTRP11; Flags: Precursor	42.4	42.4	42%	0.002	29%	Q4ZJM9.1				
<input type="checkbox"/>	RecName: Full=C1q-related factor; AltName: Full=C1q and tumor necrosis factor-related protein 14; Short=C1q/TNF-related protein 14; AltName: Full=Complement component 1 Q subcomponent-like 1; Flags: Precu	41.6	41.6	38%	0.004	31%	O75973.1				
<input type="checkbox"/>	RecName: Full=Complement C1q-like protein 4; AltName: Full=C1q and tumor necrosis factor-related protein 11; Short=C1q/TNF-related protein 11; Flags: Precursor	41.2	41.2	42%	0.004	29%	O86Z23.1				
<input type="checkbox"/>	RecName: Full=C1q-related factor; AltName: Full=C1q and tumor necrosis factor-related protein 14; Short=C1q/TNF-related protein 14; Short=CTRP14; AltName: Full=Complement component 1 Q subcomponent-lik	41.2	41.2	38%	0.004	31%	O88992.2				
<input type="checkbox"/>	RecName: Full=Complement C1q subcomponent subunit B; Flags: Precursor	40.4	40.4	49%	0.008	28%	P02746.3				
<input type="checkbox"/>	RecName: Full=Complement C1q-like protein 2; AltName: Full=C1q and tumor necrosis factor-related protein 10; Short=C1q/TNF-related protein 10; Flags: Precursor	40.4	40.4	41%	0.008	29%	Q7Z5L3.2				
<input type="checkbox"/>	RecName: Full=Complement C1q-like protein 2; AltName: Full=C1q and tumor necrosis factor-related protein 10; Short=C1q/TNF-related protein 10; Short=C1qTNF10; Short=CTRP10; Flags: Precursor	40.4	40.4	41%	0.010	29%	Q8CFR0.1				
<input type="checkbox"/>	RecName: Full=Complement C1q subcomponent subunit A; Flags: Precursor	38.5	38.5	13%	0.037	47%	P02745.2				
<input type="checkbox"/>	RecName: Full=Cerebellin-1; AltName: Full=Precerebellin; Contains: RecName: Full=Cerebellin; Short=CER; Contains: RecName: Full=[des-Ser1]-cerebellin; Flags: Precursor	37.7	37.7	28%	0.053	31%	P23435.1				
<input type="checkbox"/>	RecName: Full=Cerebellin-1; AltName: Full=Precerebellin; Contains: RecName: Full=Cerebellin; Short=CER; Contains: RecName: Full=[des-Ser1]-cerebellin; AltName: Full=des-Ser(1)-cerebellin; Short=[des-Ser1]C	37.7	37.7	28%	0.053	31%	P63182.2				
<input type="checkbox"/>	RecName: Full=Cerebellin-1; AltName: Full=Brain protein D3; AltName: Full=Precerebellin; Contains: RecName: Full=Cerebellin; Short=CER; Contains: RecName: Full=[des-Ser1]-cerebellin; Short=des-Ser1-cerebel	37.4	37.4	28%	0.071	31%	Q9R171.1				
<input type="checkbox"/>	RecName: Full=Multimerin-1; AltName: Full=EMILIN-4; AltName: Full=Elastin microfibril interface located protein 4; Short=Elastin microfibril interfacer 4; AltName: Full=Endothelial cell multimerin; Contains: RecName	36.6	36.6	25%	0.25	38%	Q13201.3				
<input type="checkbox"/>	RecName: Full=Multimerin-1; Flags: Precursor	34.3	34.3	25%	1.3	32%	B2RPV6.2				
<input type="checkbox"/>	RecName: Full=Complement C1q subcomponent subunit A; Flags: Precursor	32.3	32.3	13%	3.9	39%	Q69DL0.1				
<input type="checkbox"/>	RecName: Full=Ammonium transporter Rh type C; AltName: Full=Rhesus blood group family type C glycoprotein; Short=Rh family type C glycoprotein; Short=Rh type C glycoprotein	31.6	31.6	14%	8.9	33%	Q95JD3.1				

I risultati migliori hanno e-value = 0,001, non eccezionali (tenderebbero a zero in caso di elevatissima similarità), ma comunque significativi. L'identità tra query e subject si aggira attorno al 30% (questo perché probabilmente non ci sono sequenze di ostrica o altri molluschi depositate nel database che abbiamo consultato)

Tutti i risultati sembrano ricondurre a proteine legate a «C1q and TNF-related proteins», chiara indicazione che anche il mRNA ignoto di ostrica con ogni probabilità codifica una proteina con una funzione simile, che potrò certamente studiare grazie a dati di letteratura

STEP 2: ANNOTAZIONE STRUTTURALE

- La sequenza nucleotidica del mRNA può essere confrontata con il genoma di ostrica (se disponibile) per identificare su quale scaffold sia localizzato il gene che la codifica e da quanti esoni esso sia costituito
- Clicchiamo sul file nominato «oyster genome scaffolds»: si tratta del genoma di ostrica, che potremo utilizzare come database per un **BLAST locale** (cioè con il CLC Genomics Workbench e non più online come prima).
- Il genoma di ostrica consiste di 7659 scaffold identificate da un accession ID come evidenziato sotto, per una dimensione totale di circa 700 milioni di paia di basi

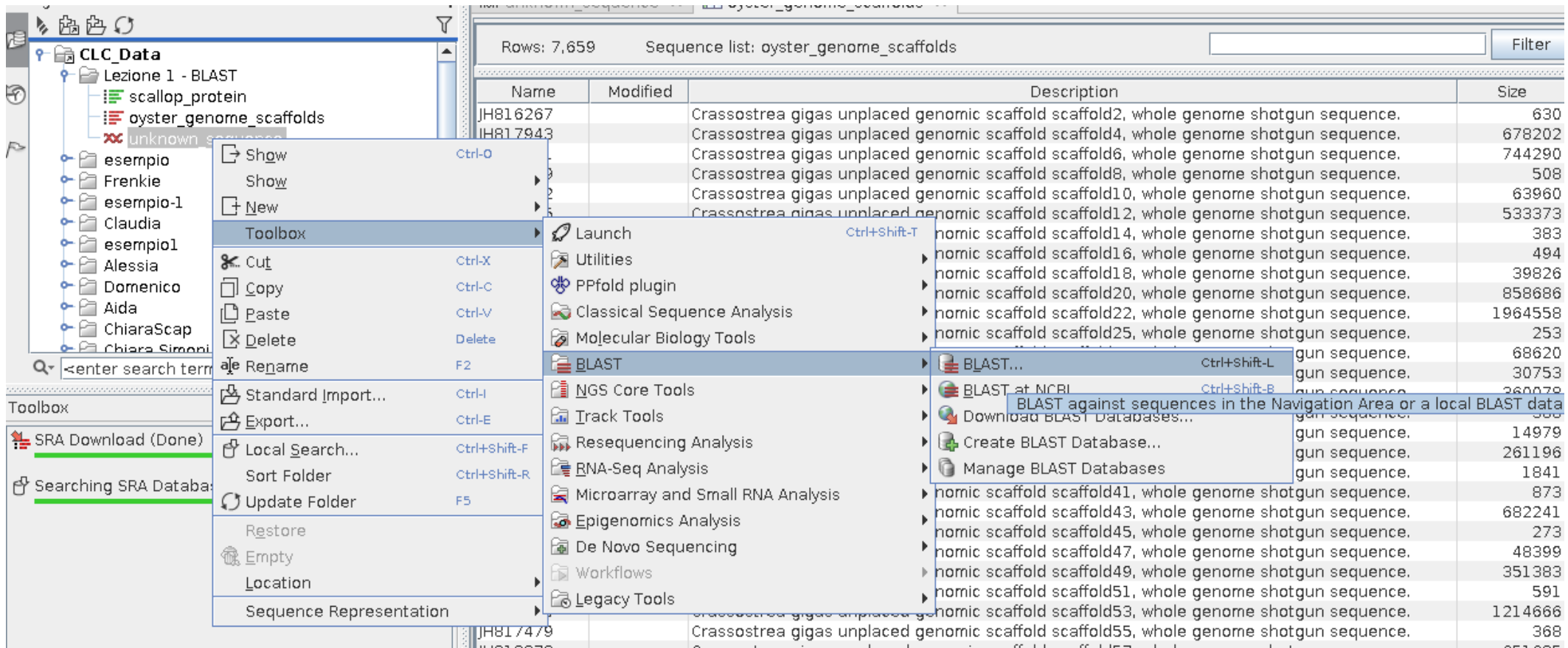


The screenshot shows the CLC Genomics Workbench interface. On the left, the 'Navigation Area' displays a tree structure under 'CLC_Data' with folders for 'Lezione 1 - BLAST', 'esempio', 'Frenkie', 'esempio-1', 'Claudia', 'esempio1', 'Alessia', 'Domenico', 'Aida', 'ChiaraScap', and 'Chiara Simoni'. The 'oyster genome scaffolds' folder is selected. The main window shows a BLAST search of an 'unknown_sequence' against the 'oyster_genome_scaffolds' database. The search results are displayed in a table with columns for scaffold ID and sequence. The scaffold JH815662 is highlighted in blue.

Accession ID	Sequence
JH816267	CTCCCCGTCTGATATATAAAATTATATCACACAAGATTTGAATACAATTTAGTACAGAAAACATTCAGTTCTCCAATAATAA
JH817943	AGAAATAATATGAACTTGCTCTTTATGACATGGGCATATTTGGTTACACACTGCTCCGAAACACCAAGGTCACATTTTAAG
JH818221	GAATATAAACTGACATGTGCTGCGGTGTTTCTGTTTTATTTAGTGCCATCGATAAATGTTTCAGAAACTTTTGGTGATTGCACC
JH818459	ACTCGCACTCATCCCTGGGGCAGGGGTACATTGTCACTTTTCAGTACTCCGGCACCACGGAGTGTCACAGTTGATAGTCAAG
JH815662	TTATTTCAAGTTACATTATTAATATTAGGCAATCTCAAATTTGGAAAAACACACTCAAATTTGGGAAAAAGATGCCTTTTTT
JH816405	TCAGATAAAACAATTTGCTTTACACAGGAAAAC TAGCATTATCAAATATATGCATTTACATGTGTCTATATTTAAAAAAC
JH816996	TCTATATGCTTCATTTGGCGCTACGGTGATTTGATATGACCTTTCTGACGTGAATGTGCAGATTGCACATAAATCGGAAT

Selezioniamo la sequenza ignota con il tasto destro e lanciamo un BLAST selezionando Toolbox -> BLAST -> BLAST come mostrato sotto

Vogliamo confrontare la sequenza del mRNA con il genoma per trovare su quale scaffold è localizzato il gene e trovare la localizzazione precisa degli esoni



Dobbiamo selezionare BLASTn dal menù a tendina (confronto di sequenza nucleotidica contro database nucleotidico)

Inoltre bisogna selezionare il database «oyster_genome_scaffolds» sotto «target» cliccando sull'icona indicata dal cerchio

Navigation Area

- CLC_Data
 - Lezione 1 - BLAST
 - scallop_protein
 - oyster_genome_scaffolds
 - unknown_sequence
 - esempio
 - Frenkie
 - esempio-1
 - Claudia
 - esempio1
 - Alessia
 - Domenico
 - Aida
 - ChiaraScap
 - Chiara Simoni

Toolbox

SRA Download (Done)	100 %
Searching SRA Database...	100 %

BLAST@labgenCLC

- Select sequences of same type
- Choose program and target**
- Set BLAST parameters
- Result handling

Choose program and target

BLAST program

Program: blastn: DNA sequence and database

Target

Sequences

oyster_genome_scaffolds

BLAST database

FATTO QUESTO, CLICCHIAMO SU «NEXT»

Help Reset Previous **Next** Finish Cancel

BLAST@labgenCLC

1. Select sequences of same type
2. Choose program and target
3. **Set BLAST parameters**
4. *Result handling*

Set BLAST parameters

Choose parameters

Number of threads: 24

Choose filter: Filter low Complexity

Expect: 10.0

Word size: 32

Match/mismatch: Match 2, Mismatch -3

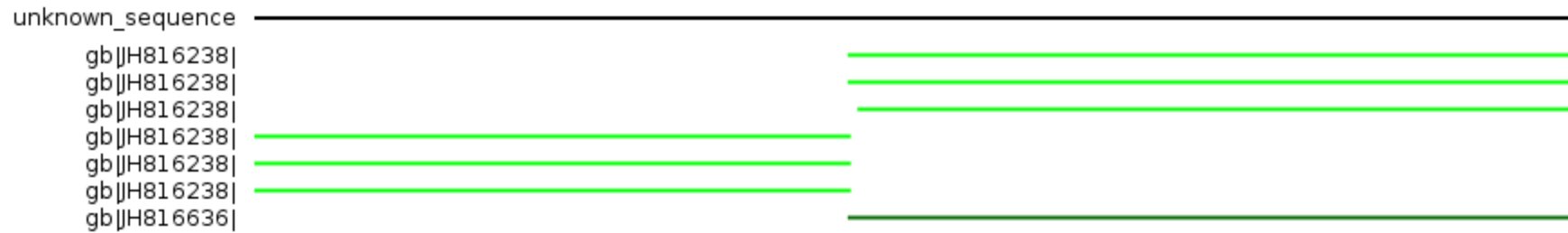
Gap costs: Existence 5, Extension 2

Max number of hit sequences: 250

Help Reset Previous Next Finish Cancel

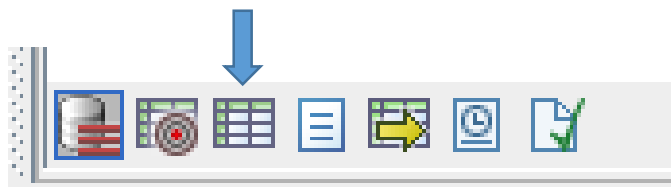
Lasciamo questi parametri (il cui significato è stato accennato brevemente a lezione) con i valori di default

Un word size = 32 andrà bene in questo caso perché stiamo cercando una sequenza che deve essere identica al mRNA, quindi ci basta una bassa sensibilità, ma questo ci aiuterà a rendere la ricerca molto veloce



Osserviamo i risultati... la parte grafica è simile a quella che abbiamo visto online, anche se va dal verde chiaro (molto significativo) al nero (poco significativo)

- 1) Gli hit di BLAST sono divisi in due sezioni rispetto alla nostra sequenza ignota, probabile indicazione che l'mRNA è codificato da un gene con due esoni
- 2) Tutti gli hit altamente significativi sono localizzati sulla stessa scaffold (JH816238), che possiamo andare a recuperare dal file «oyster_genome_scaffold» effettuando una ricerca testuale
- 3) Come mai troviamo tre hit con la stessa significatività sia per l'esone 1 che per l'esone 2? Questa potrebbe essere un'indicazione che il nostro gene di interesse è presente in 3 copie identiche (geni paraloghi) sulla stessa scaffold



Andiamo a vedere il dettaglio dei risultati dall'output tabulare cliccando sul simbolo evidenziato a fianco

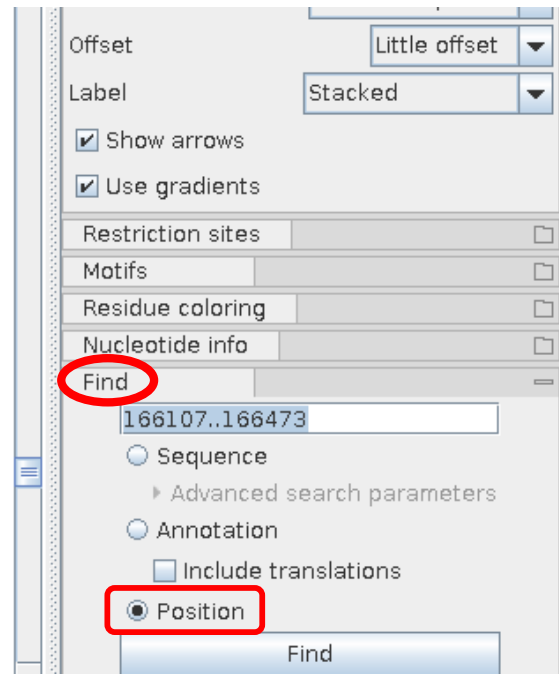
Rows: 7 Summary of HSPs from query: unknown_sequence <input type="text"/> Filter ▾							
Hit	E-value	HSP start	HSP end	Query start	Query end	%Identity	HSP frame/stra...
JH816238	0.00	165866	165416	366	816	100.00	Minus
JH816238	0.00	194947	194497	366	816	100.00	Minus
JH816238	0.00	188921	188478	372	816	99.10	Minus
JH816238	0.00	166473	166107	1	367	100.00	Minus
JH816238	0.00	195554	195188	1	367	100.00	Minus
JH816238	0.00	190208	189842	1	367	99.46	Minus

Per comodità sono mostrate soltanto alcune delle varie colonne che possono essere selezionate. Notiamo che:

- 1) Tutti gli hit positivi sono sullo strand Minus (ultima colonna), HSP = High Scoring Pairs, cioè la regione di hit trovata nella scaffold. Questo significa che **tutti e tre i geni sono localizzati sullo strand reverse** e che quindi dovranno essere annotati da destra verso sinistra.
- 2) Query start/end indicano la posizione di hit nel mRNA, HSP start/end la posizione dove inizia e finisce l'hit nella scfold. Ovviamente posso supporre che l'esone 1 e 2 più vicini facciano parte dello stesso gene
- 3) Quindi il gene 1 comprenderà: esone 1 (166107-166473) + esone 2 (165416-165866)
- 4) Il gene due sarà distante circa 23 Kb ed inizierà in posizione 188478
- 5) Il terzo gene sarà più vicini al secondo (circa 4 Kb), iniziando in posizione 194497

Posso aprire la sequenza della scaffold di interesse con un doppio click e annotare a mano i due esoni del gene 1, come mostrato sotto

```
JH816238 TGAATAAATACTATATTAATGGTATATTGGCCTCAATTGAAATTTAACATCATAAATAAAAGGTTGATTCTTTTAAACAA
166,100 166,120 166,140 166,160
JH816238 TCAGATATATATTTTCATTAGCTTACCTTTTTCAGGAATTCAGATCGTTTAAAAGCCGAAGCTTCTTCATCTGGACAT
166,180 166,200 166,220 166,240
JH816238 TCATTGATGTGGGTTTCGTTTAAATATTGATGAATTTTAAATATTAGTTTACATGCGCGTTGGCGTTGCAAGTGCAATATT
166,260 166,280 166,300 166,320
JH816238 TTTTCATCTTGAGTTTGAATCATAGTTCGAAACTCTGCTATTTCTTCAGATTGCATAGCATTGTCTCTCTAATGCCAG
166,340 166,360 166,380 166,400
JH816238 TATTCATTTTATTTGATCCAAAAGTAAAGATTTAGCTCAAAAATCATATTTTCGTAATTTGATTCATCCTTTACATGTA
166,420 166,440 166,460 166,480
JH816238 TTTTTCCTCAGTTTGGCAGAACTAATTGAAAAATATCCAAAAGAAGCAAACAATTGCGGGCAAAATAAATGCAATGTCCTT
166,500 166,520 166,540 166,560
JH816238 TCTTGCAGTGACAAGCGACAAACTGTGTCCTTATCAATTACAATTCAAGTTTATTTCGATACAATGACATTTAAAGGGAG
166,580 166,600 166,620 166,640
JH816238 AAAGACGTTTACCACGTGATCTTTCAGAATACATGTTTAAACATTAATAATTAACGAGAGAGAGAGAGAGAGAGANN
166,660 166,680 166,700 166,720
JH816238 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
166,740 166,760 166,780 166,800
JH816238 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```



Dal menù a destra cerchiamo «Find», quindi selezioniamo «position»

Scriviamo a mano le coordinate dell'esone 1 (166107 e 166473, cioè i nucleotidi iniziali e finali di hit), attenzione a mettere prima il numero più basso

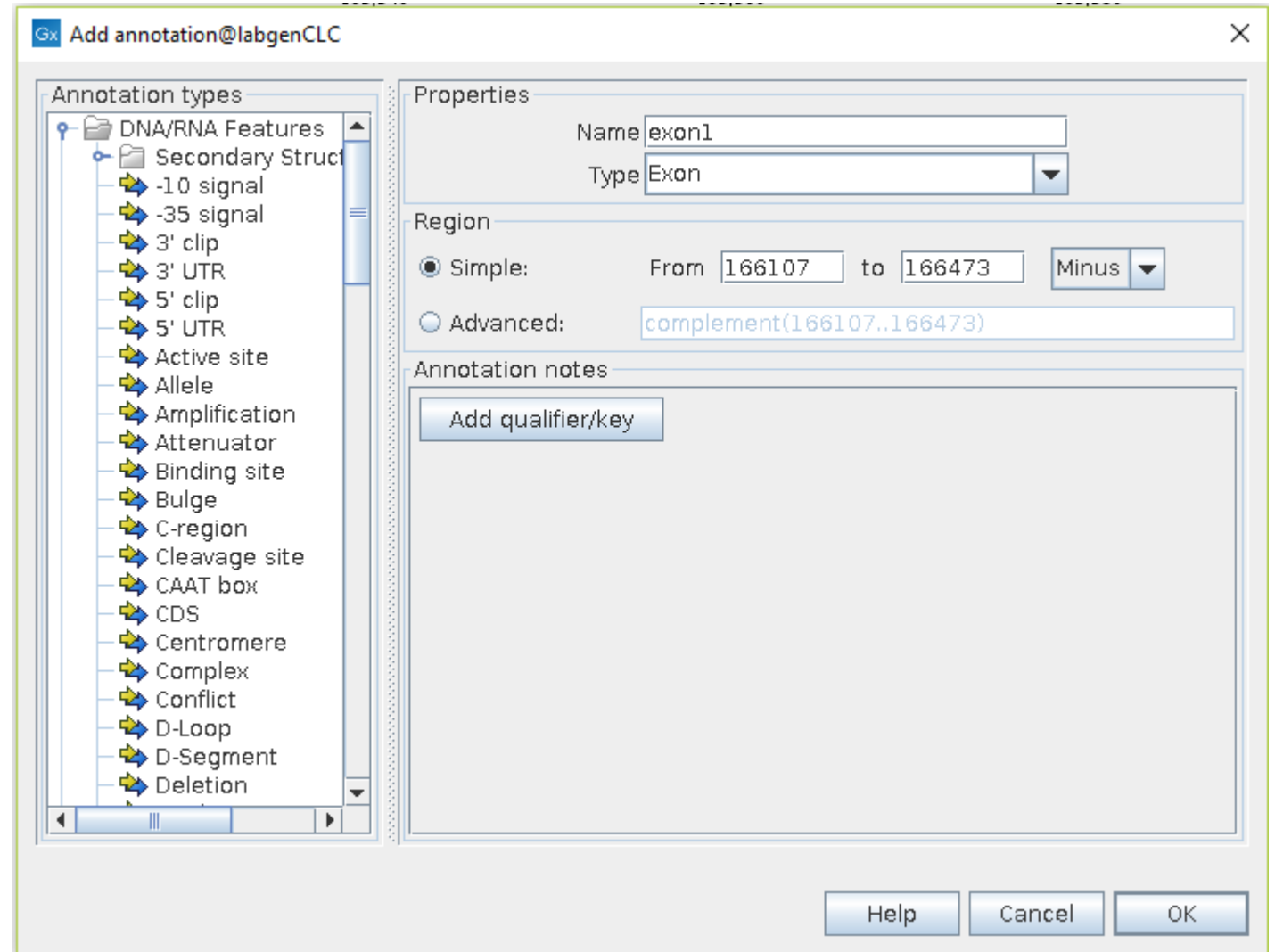
Clicchiamo su «Find» e la regione verrà automaticamente evidenziata

Clicchiamo sulla selezione con il tasto destro e selezioniamo «add annotation»

Dal menù a tendina su «Type» scegliamo exon

Come «name» indichiamo «exon1»

Ricordiamoci di selezionare «Minus per quanto riguarda lo strand!»



Possiamo ripetere la stessa operazione per il secondo esone e, volendo, fare lo stesso per gli altri due geni. Allo stesso modo potremmo inserire le annotazioni degli introni o altri elementi di regolazione

STEP 3: ANALISI COMPARATA

- Quanti geni codificanti proteine appartenenti a questa famiglia sono presenti in un altro mollusco bivalve?
- Verifichiamolo con un BLAST nella specie *Mizuhopecten yessoensis*
- Questo ci permetterà di verificare, grossolanamente, anche quanto le due specie siano divergenti tra loro
- Il proteoma completo di questa specie è contenuto nel file «scallop_protein», che potremo utilizzare come database per un altro BLAST locale, partendo dalla nostra sequenza ignota come query
- Attenzione al tipo di BLAST in questo caso, perché la combinazione è tra query nucleotidica e database proteico, quindi ci servirà un BLASTx


1. Select sequences of same type
2. **Choose program and target**
3. Set BLAST parameters
4. Result handling

Choose program and target

BLAST program

Program: blastx: Translated DNA sequence and protein database ▾

Target

 Sequencesscallop_protein  BLAST database

Questi i settings da
selezionare...

Help

Reset

Previous

Next

Finish

Cancel

Rows: 126 Summary of HSPs from query: unknown_sequence Filter

Hit	E-value	%Identity
PY_T14461	1.99E-17	25.90
PY_T16129	8.36E-17	37.37
PY_T19636	2.45E-15	27.53
PY_T24515	1.66E-14	36.25
PY_T20648	3.09E-14	35.04
PY_T19634	3.09E-14	27.98
PY_T22932	2.33E-13	29.29
PY_T18142	2.34E-13	38.66
PY_T22931	2.21E-12	30.77
PY_T16846	7.84E-12	38.24
PY_T14104	1.25E-11	30.15
PY_T16848	1.77E-11	42.68
PY_T14101	3.25E-11	34.93
PY_T14616	6.48E-11	26.81
PY_T22690	1.25E-10	27.57
PY_T19635	1.26E-10	22.27
PY_T19483	1.46E-10	28.16
PY_T16771	2.44E-10	35.00
PY_T21371	2.53E-10	33.06
PY_T21373	2.83E-10	33.02
PY_T16803	3.37E-10	29.77
PY_T16542	6.06E-10	26.72
PY_T22522	1.66E-9	31.30
PY_T24690	2.22E-9	30.71
PY_T14102	3.05E-9	35.42
PY_T14015	3.48E-9	43.28
PY_T04594	3.53E-9	35.71
PY_T22521	4.55E-9	28.68
PY_T23591	5.12E-9	25.35
PY_T24684	5.26E-9	24.36
PY_T19330	6.78E-9	35.37
PY_T18563	1.44E-8	31.08
PY_T00976	1.44E-8	34.29
PY_T04595	1.55E-8	22.01
PY_T22689	2.34E-8	31.30

L'e-value migliore è 1.99×10^{-17} , certamente significativo ma comunque abbastanza lontano da zero

Interessante è la % di identità, che ci dice che la proteina di *M. yessoensis* che più assomiglia a quella di ostrica è identica soltanto per il 25,9%!

Possimo concludere che la distanza evolutiva tra le due specie sia considerevole, forse più di quanto ci aspettassimo

ALIGNMENTS

>PY_T14461 No definition line
Length=281

Score = 80.1 bits (196), Expect = 2E-17
Identities = 65/251 (25%), Positives = 115/251 (46%), Gaps = 26/251 (10%)
Frame = +1

```
Query 115 ELKSLLLDQNERILALERQNAMQSEETAEFRMTMIQTQDEKILHLQ-----R 252
+++ LL +I +L+R+NA E+ I T +EK+ L+
Sbjct 37 DIRLLLLNKYGSQIESLQRENARLEEKMTLEKKIMTLEEKVTTLEDQTPAFEDKTPMFEE 96

Query 253 QRACKTNIKNSS-----ILNEPTSMNVQMKKASAFKRSGIPEKVIQTSAQQRIRARAADTT 417
QR K+++ N++ IL P + + K+ + KR+ P++ ++ R+ D+
Sbjct 97 QREPKSDLVNMTKPGPDILGIPDGLAISDGKSQSRKRN--PDEAVK-KVNPVRVDSTTDSI 153

Query 418 IAFYAYSSHTFPRPSDHFILNFDTIITNLGNGYHPHSGTFIATRSGYYVFTWSFRIQGDA 597
IAF+A SHT P+ I+ FD IITN+G Y +G F G Y F+W +
Sbjct 154 IAFHAILSHTVTDPAAEHIVTFDHIITNIGAHYSSKTGVFTCVEVGVYQFSWMIEVASVQ 213

Query 598 HISTELVVNDTPRGTVYYDATEKXXXXXXXXXXXXXXXXQDDEVFIRTTNTNYNLGDILSDN 777
I+T+LV N G+ + + + + DEV++R T +++G ++
Sbjct 214 WITTQLVRNGAVIGSA-ISGDDTYWTTGAASAITMLAPGDEWVWRLA-TGHSVGDITPT 271

Query 778 AGRTYFGGALL 810
T F G+L+
Sbjct 272 --YTMFNGFLI 280
```

Nel dettaglio osservate l'allineamento tra le due proteine

Sebbene la similarità non sia particolarmente elevata, la lunghezza dell'allineamento è sufficiente per portare l'e-value a livelli di significatività piuttosto elevati

Ricordate che l'e-value dipende sia dalla similarità che dalla lunghezza dell'allineamento