

RICERCARE LE SEQUENZE AMINOACIDICHE DELL'INSULINA

Andare sul sito: <http://www.uniprot.org/>

In alto compare una barra di ricerca nella banca dati di Uniprot (Universal Protein Resource): questo è il catalogo mondiale più completo per quanto riguarda le informazioni sulle proteine. È un'archivio centrale di sequenze proteiche, creato unendo le informazioni contenute in Swiss-Prot, TrEMBL (ossia *Translated EMBL*, il database ottenuto dalla traduzione delle informazioni genetiche contenute nel database dell'EMBL) e PIR (Protein Information Resource, una risorsa bioinformatica pubblica e integrata, per la ricerca genomica e proteomica, presso il Georgetown University Medical Center-GUMC).

Fare clic su "Fields". Selezionare "Taxonomy" e scrivere "chordata" nel campo "Term" che compare selezionare.

Fare clic su "Add and search" e poi di nuovo su "fields"; questa volta selezionare "Gene name" e scrivere "ins".

Nella lista dei risultati selezionare:

INS_HUMAN, INS_BOVIN, INS_PIG, INS_CHICK, INS_MYXGL, INS_ONCKE, INS_HORSE, INS_HYDCO, INS_PANTR, INS_RABIT, INS_SHEEP, INS1_MOUSE, INS2_XENLA

All	Accession	Entry name	Status	Protein names	Gene names	Organism
<input checked="" type="checkbox"/>	P01308	INS_HUMAN	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Homo sapiens (Human)
<input checked="" type="checkbox"/>	P01317	INS_BOVIN	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Bos taurus (Bovine)
<input checked="" type="checkbox"/>	P01315	INS_PIG	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Sus scrofa (Pig)
<input type="checkbox"/>	P01321	INS_CANFA	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Canis familiaris (Dog)
<input type="checkbox"/>	P01329	INS_CAVPO	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Cavia porcellus (Guinea pig)
<input checked="" type="checkbox"/>	P67970	INS_CHICK	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Gallus gallus (Chicken)
<input checked="" type="checkbox"/>	P01342	INS_MYXGL	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	ins	Myxine glutinosa (Atlantic hagfish)
<input checked="" type="checkbox"/>	P04667	INS_ONCKE	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	ins	Oncorhynchus keta (Chum salmon)
<input type="checkbox"/>	P12703	INS_ALLMI	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	INS	Alligator mississippiensis (American alligator)
<input type="checkbox"/>	P29335	INS_AMICA	★	Insulin [Cleaved into: Insulin B chain; Insulin A chain]	ins	Amia calva (Bowfin)

Nella barra verde in basso, fare clic su "Retrieve" e nella nuova schermata selezionare "Download" nel formato FASTA. Il file con le sequenze è così salvato su computer con un nome dato dal programma (ricordarsi del nome).

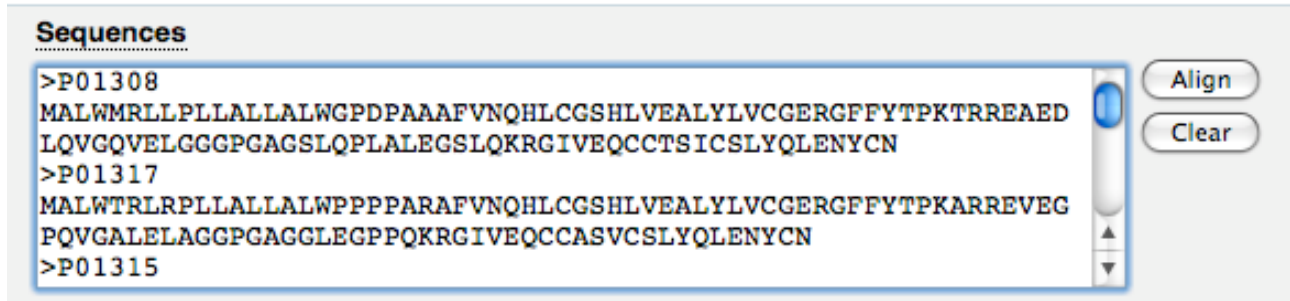
Download data [compressed](#) or [uncompressed](#)

FASTA
Sequence data in FASTA format.
[\[Download \(6 KB\) | Open \]](#)

CONFRONTARE LE SEQUENZE AMINOACIDICHE DELL'INSULINA E COSTRUIRE UN ALBERO FILOGENETICO

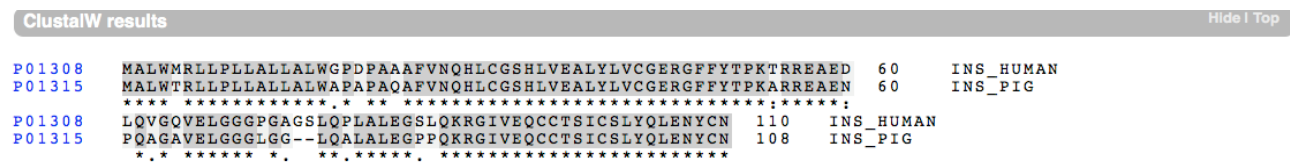
Nella barra verde in basso fare clic su "Align". In questo modo il software fa un allineamento multiplo per comparare le sequenze amminoacidi che scelte.

Nella nuova schermata, in alto si possono scorrere tutte le sequenze amminoacidi che dell'insulina dei diversi animali selezionati.

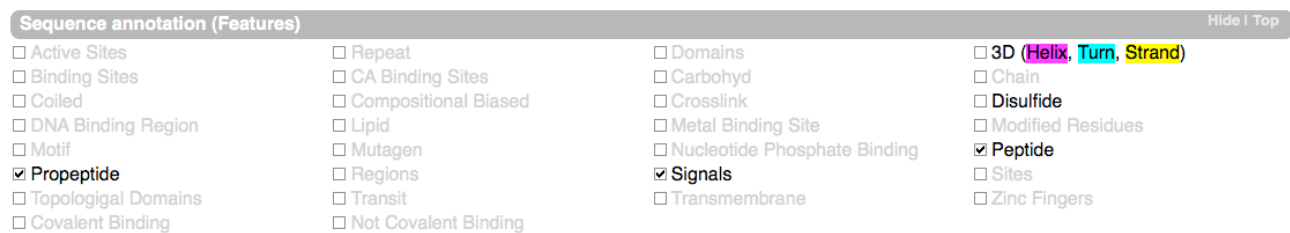


Prima di confrontare contemporaneamente tutte le 13 sequenze, selezionare nella lista sottostante solo INS_HUMAN e INS_PIG. Fare di nuovo clic su "Align" nella barra verde in basso.

Compare il confronto tra le due sequenze. I trattini orizzontali " - " indicano un gap nell'allineamento. Gli amminoacidi evidenziati in grigio scuro sono quelli conservati nelle sequenze delle due specie: il simbolo utilizzato è un asterisco (*). In grigio più chiaro con simbolo " : " sono evidenziate differenze che non influenzano la struttura dell'insulina. Sempre in grigio ma con simbolo " . " sono indicate differenze di amminoacidi che hanno effetto sulla struttura della molecola. Infine dove non vi è alcun simbolo sono indicati amminoacidi differenti in ciascuna specie.



Fare clic su "Show" nella sezione "Sequence annotation" e selezionare "signals", "propeptide" e "peptide" per evidenziare il peptide segnale (verde), il peptide C (rosso), i peptidi A e B (viola).



Per evidenziare invece le sequenze formanti alfa eliche, beta foglietti e piegamenti, selezionare "3D": scegliere invece "Disulfide" per ottenere gli amminoacidi che formano ponti disolfuro.

- 1) Qual è la percentuale di conservazione nei diversi segmenti? Completa la tabella:

	Peptide segnale		Catena B		Peptide C		Catena A	
	n° *	%	n° *	%	n° *	%	n° *	%
INS_HUMAN VS INS_PIG								

- 2) Quali sequenze sono maggiormente conservate? Quali meno? Che considerazioni puoi fare?
- 3) Perché pensi siano stati inseriti dei trattini (-) nell'allineamento della sequenza INS_CHICK? Quale può essere l'interpretazione dal punto di vista evolutivo?

Tornando indietro si possono selezionare le sequenze INS_HUMAN e INS_PANTR in modo da poterle confrontare.

Quindi si può provare a confrontare fra di loro tutte le sequenze aminoacidiche di insulina a disposizione. Tornare indietro e selezionare tutte le sequenze, quindi di nuovo far clic su “Align”.

Osservare i risultati, selezionando, come prima, i diversi tratti delle molecole.

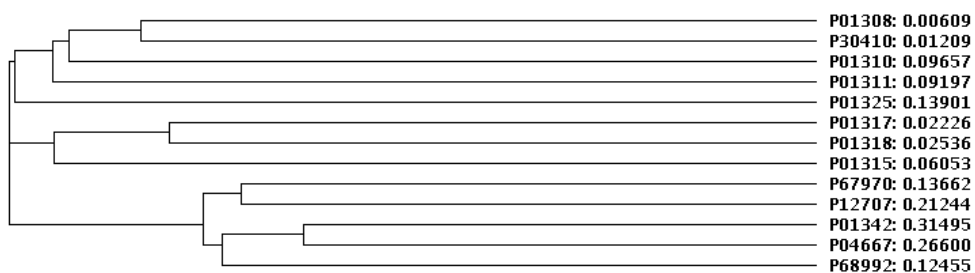
ClustalW results

P67970	MAL--WIRSLPLLALLVFSGPGTSYAAANQHLCGSHLVEALYLVCGERGFFYSPKARRDV	58	INS_CHICK
P12707	MAL--WMQCLPLVLVLLFSTPNT--EALANQHLCGSHLVEALYLVCGERGFFYYPKIKRDI	57	INS2_XENLA
P01308	MAL--WMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREA	58	INS_HUMAN
P30410	MAL--WMRLLPLLALLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREA	58	INS_PANTR
P01310	-----FVNQHLCGSHLVEALYLVCGERGFFYTPKAXEA	34	INS_HORSE
P01311	MAS--LAALPLLALLVLCRLDPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKSRREV	58	INS_RABIT
P01325	MAL--LVHFLPLLALLALWEKPKTQAFVKQHLGPHLVEALYLVCGERGFFYTPKSRREV	58	INS1_MOUSE
P01317	MAL--WTRLRPLLALLALWPPPARAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREV	58	INS_BOVIN
P01318	MAL--WTRLVPLLALLALWAPAPAHAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREV	58	INS_SHEEP
P01315	MAL--WTRLRPLLALLALWAPAPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREA	58	INS_PIG
P68992	-----VPTQRLCGSHLVDALYFVCGERGFFYSPKPIREL	34	INS_HYDCO
P04667	MAF--WLQAASLLVLLALS--PGVDAAAAQHLCGSHLVDALYVCGEKGFYTPK--RDV	54	INS_ONCKE
P01342	MALSPFLAAVILVLLLSRAPPSADTRTTGHLCKGDLVNALYIACGVRGFFYDPTKMKRD	60	INS_MYXGL
.:***:.* :**** *			
P67970	EQPLVSSPLRGEAG---VLPFQQE---EYEKVKGIVEQCCHNTCSLYQLENYCN	107	INS_CHICK
P12707	EQAQVNGPQDNELD---GMQFQPQ---EYQMKRGIVEQCCHSTCSLFQLENYCN	106	INS2_XENLA
P01308	EDLQVGQVELGGGPGAGSLQPLAL---EGSLQKRGIVEQCCTSICSLYQLENYCN	110	INS_HUMAN
P30410	EDLQVGQVELGGGPGAGSLQPLAL---EGSLQKRGIVEQCCTSICSLYQLENYCN	110	INS_PANTR
P01310	EDPQVGEVELGGGPGGLQPLAL---AGPQQXGIVEQCCTGICSLYQLENYCN	86	INS_HORSE
P01311	EELQVGQAEELGGGPGAGGLQPSAL---ELALQKRGIVEQCCTSICSLYQLENYCN	110	INS_RABIT
P01325	EDPQVEQLELGGSP--GDLQTLAL---EVARQKRGIVDQCCTSICSLYQLENYCN	108	INS1_MOUSE
P01317	EGPQVGALELAGGP--G---AGGL---EGPPQKRGIVEQCCASVCSLYQLENYCN	105	INS_BOVIN
P01318	EGPQVGALELAGGP--G---AGGL---EGPPQKRGIVEQCCAGVCSLYQLENYCN	105	INS_SHEEP
P01315	ENPQAGAVELGGGL--GGLQALAL---EGPPQKRGIVEQCCTSICSLYQLENYCN	108	INS_PIG
P68992	EP-----LLGIVEQCCHNTCSLANLEGYCN	59	INS_HYDCO
P04667	DP--LIGFLSPKSAKENEYFPKQD---TEMMVKRGIVEQCCHKPCNIFDLQNYCN	105	INS_ONCKE
P01342	TGALAAFLPLAYAEDNESQDDESIGINEVLKSREGIVEQCCHKRCISYDLENYCN	115	INS_MYXGL
: *.:*:.***			

- 1) In diverse regioni delle sequenze non vi è conservazione di aminoacidi tra le varie specie. A quali segmenti corrispondono?
- 2) In quali tratti della catena polipeptidica invece compaiono gli asterischi?
- 3) Che considerazioni puoi fare a riguardo?

Nella sezione “ClustalWtree” compare l’albero filogenetico costruito in base alle somiglianze tra le sequenze. In un albero filogenetico si distinguono “rami”, che si sviluppano a partire da “nodi” (un punto di divergenza evolutiva da un antenato comune) e che terminano in “foglie”, che corrispondono alle sequenze presenti nei singoli taxa. La lunghezza dei “rami” indica il tempo trascorso da un evento di divergenza. In particolare quello rappresentato è un albero “con radici” ossia si assume un evento, evolutivamente molto distante rispetto agli altri, come punto di partenza (la “radice”).

ClustalW tree



Dopo aver associato a ogni sigla il nome dell’animale, prova a rispondere.

- 1) Quale specie è più affine alla mucca? Quale più affine al salmone?
- 2) Osserva le coppie “scimpanzè-uomo”, “salmone-missina”, “pecora-mucca”. Si tratta coppie di animali differenziate ciascuna da un unico “nodo”. Quali differenze puoi notare nelle divergenze di queste coppie?
- 4) Basandoti sulla similarità di sequenza dell’insulina, quale specie animale potrebbe fornire insulina “per uso umano”?
- 5) Pensi che l’utilizzo di questo animale, possa avere implicazioni etiche e commerciali (considera se si tratta di una specie diffusa, facilmente allevabile, che non sia protetta, etc.)?
- 6) Sulla base di queste considerazioni, e tenendo conto anche della similarità di sequenza, quale animale potrebbe allora essere un “donatore” ideale?

CLUSTALW è un software disponibile in più siti di bioinformatica: ad esempio proviamo ad aprire il sito <http://align.genome.jp/> . Per visualizzare un albero filogenetico fare clic su “Sfoglia” e selezionare il file salvato, in formato FASTA, con le sequenze dell’insulina. Quindi fare clic su “Execute Multiple Alignment”.



Multiple Sequence Alignment by CLUSTALW

CLUSTALW
MAFFT
PRRN

[Help](#)

General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment: FAST/APPROXIMATE SLOW/ACCURATE

Enter your **sequences** (with labels) below (copy & paste): PROTEIN DNA

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

Or give the file name containing your query

Sfoglia...

Execute Multiple Alignment
Reset

Nella nuova schermata, scorrendo verso il basso, si ritrova l’allineamento multiplo delle sequenze e, in fondo, si può scegliere il formato dell’albero filogenetico. Scegliere “Dendrogram with branch length” e quindi “Exec”: si ottiene così l’albero filogenetico.

