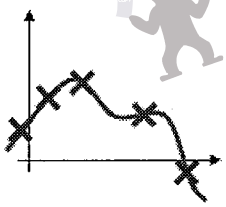
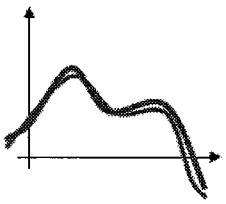
Bioinformatica

Introduzione

Per studiare un fenomeno, i biologi si concentrano sulla parte sperimentale, ovvero cercano di vedere come si collocano i punti sulla curva relativa ad un certo fenomeno, per poi classificarli. Gli informatici invece di focalizzarsi sul singolo punto, realizzano dei modelli, cioè una forma che meglio ripercorra quella curva y=f(x). Di fatto cercano un'approssimazione.

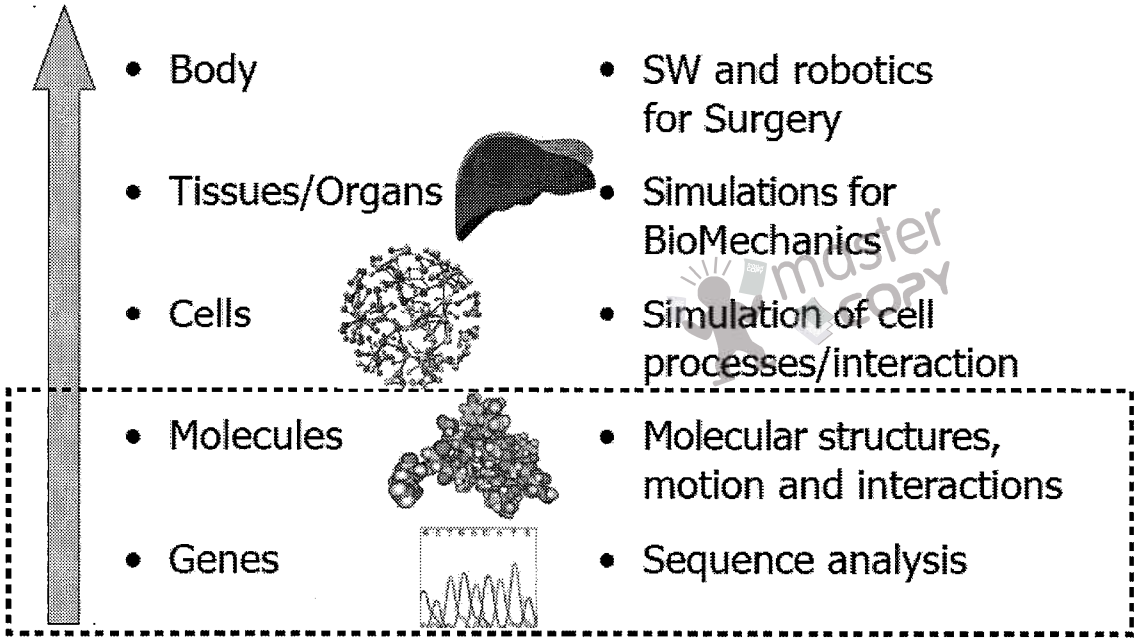


$$y = f(x)$$



Le due prospettive sono complementari, nessuna prescinde dall'altra. Per risolvere problemi biologici complessi verranno sfruttate entrambe.

I problemi risolvibili con questo approccio sono molti: si parte dai geni con l'analisi di sequenza per arrivare a livello molecolare. Si passa poi a livello cellulare dove vengono realizzate simulazioni delle interazioni tra cellule; a livello tissutale invece i problemi diventano di tipo biomeccanico. Infine, a livello di organismo si può lavorare con software associati di robotica e diagnostica.

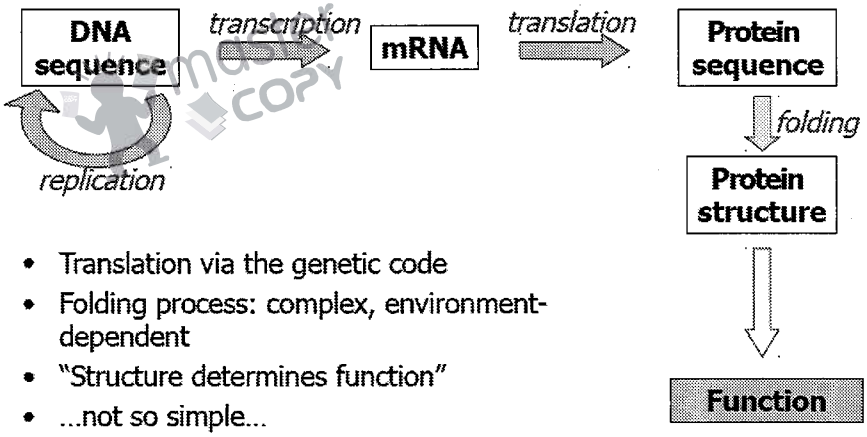


Il corso terrà in considerazione solo analisi di sequenza e interazioni tra molecole.

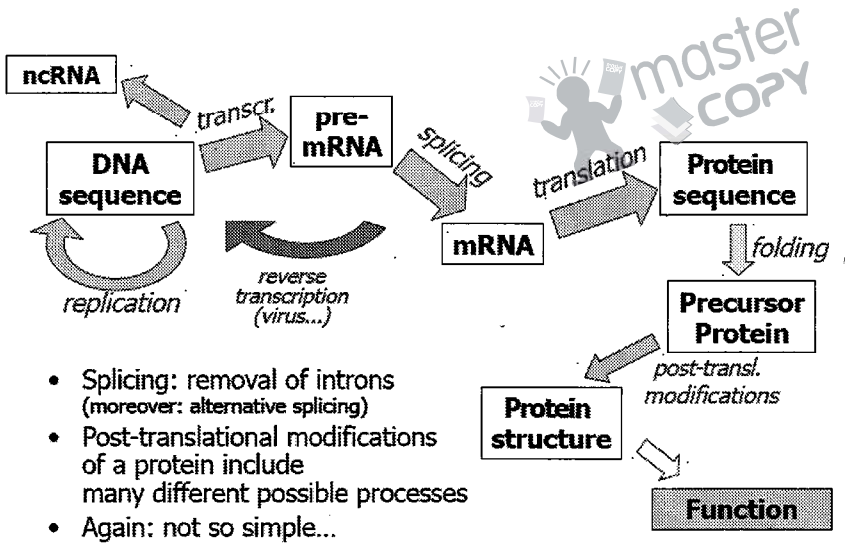
Dogma centrale della biologia molecolare

Le informazioni che regolano l'intero organismo sono contenute nel DNA. Queste informazioni però per essere sfruttate devono seguire una serie di passaggi.

L'informazione del DNA viene utilizzata per generare mRNA o RNA messaggero tramite trascrizione. Da qui si ha la sintesi di una sequenza proteica, che poi in seguito a ripiegamento darà vita ad una proteina, la quale conterrà tutte le funzionalità scritte nel DNA.



In realtà in questo modo il quadro risulta semplificato. Infatti, innanzitutto il passaggio da DNA e mRNA non è diretto ma c'è un processo che richiede che alcune sequenze vengano rimosse. Si parla di SPLICING. Questo è dovuto al fatto che non tutta la sequenza di DNA è codificante. Lo stesso splicing per altro può avvenire secondo diverse modalità.
Inoltre, può accadere che dall'mRNA si torni al DNA. Poi non tutto l'mRNA prodotto viene utilizzato per la sintesi proteica, ma anche per altre funzionalità.
Infine, dalla sequenza proteica si possono verificare altre modifiche prima di arrivare alla proteina, e così via...

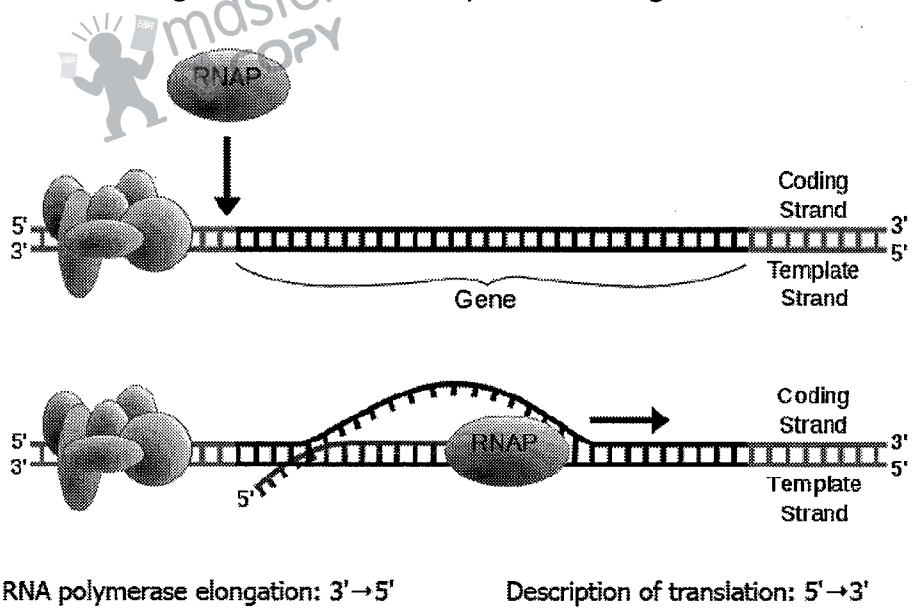


Richiami di biologia molecolare

Trascrizione RNA

Si parte con un insieme di proteine che inducono la trascrizione, la quale avviene attraverso l'RNA polimerasi. Ogni filamento di DNA è caratterizzato tra una direzionalità 3'->5' e 5'->3'.

A partire dai promotori l'RNA polimerasi identifica un gene, si lega e scorre lungo il filamento di stampo da 3' a 5'. Per ogni base, aggiunge al filamento di RNA di accrescimento, la base corrispondente. Il filamento generato è identico a quello di coding del DNA.



RNA polymerase elongation: 3'→5' Description of translation: 5'→3'

Genoma: insieme di tutto il materiale genetico di un organismo. Non è detto che tutto il materiale sia codificante, cioè soggetto a trascrizione.

Trascrittoma: insieme di tutte le sequenze di RNA trascritte dal genoma. A seconda della cellula la parte trascrivibile cambia ed è ciò che è alla base della differenziazione tra i tessuti.

Esoma: insieme di tutte le sequenze codificanti (CDS)

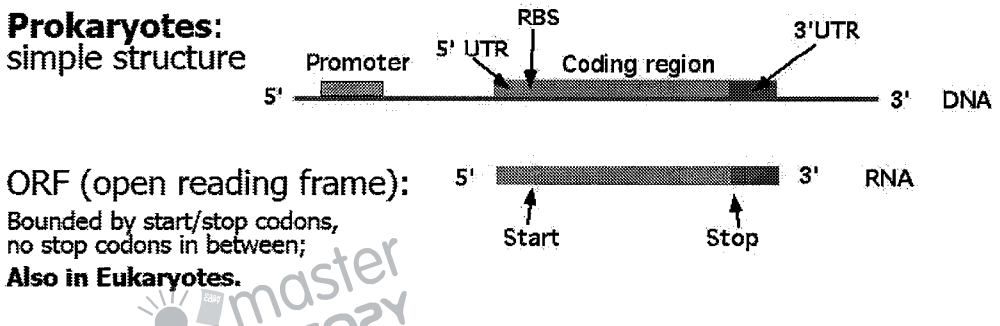
Struttura di un gene

Nel genoma si hanno porzioni che partecipano nel codificare una proteina. Si parla di **geni**. La struttura del genoma è molto diversa tra eucarioti e procarioti.

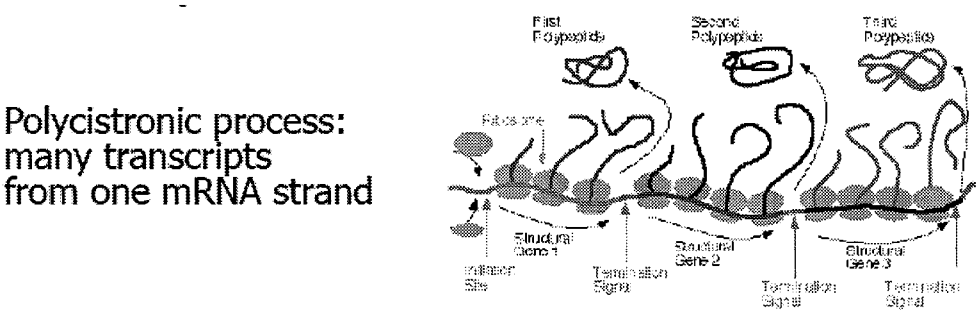
Procarioti

Dato il filamento del DNA si identifica una porzione nel lato 5' chiamato ***promotore***, dove si legano gli enzimi che danno inizio alla trascrizione. A sinistra e a destra della regione codificante ci sono due regioni che vengono trascritte ma non contengono informazioni sulla proteina. La regione

interessante è quella centrale, chiamata **OPEN READING FRAME**. Essa è delimitata da un ***codone di start*** e uno di ***stop***. Ogni codone è formato da 3 paia di basi ed è importante che al centro non ci sia alcun codone di stop.

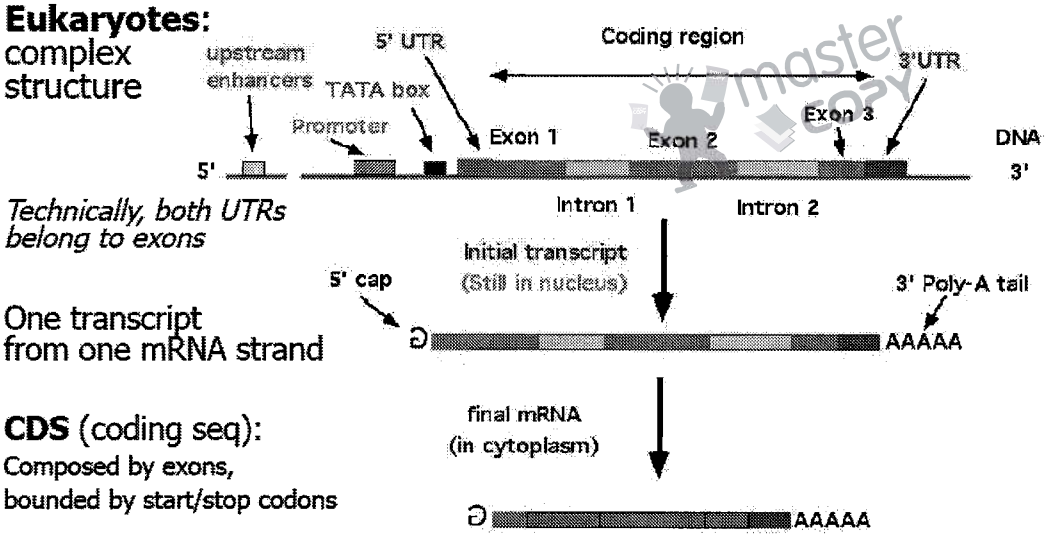


Nei procarioti la sintesi proteica prevede che si trascriva un filamento che contiene diversi geni. Sullo stesso filamento vanno ad agire più ribosomi. Il ribosoma e il gene scorrono l'uno rispetto all'altro generando filamenti di polipeptide. Si parla di **processo policistronico**.



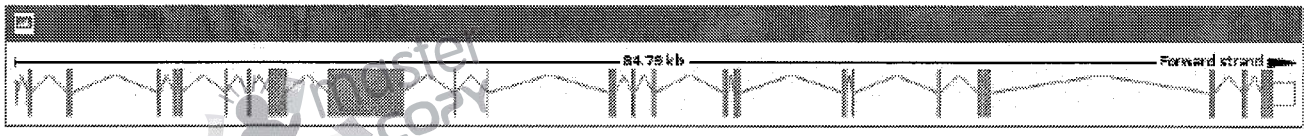
Eucarioti

In questo caso a monte del promotore ci sono altri fattori. In particolare, tra il promotore e la parte codificante c'è una zona con una ripetizione di TATA chiamata appunto TATA BOX, che controlla il processo, e poi c'è una zona che viene trascritta intervallata da due UTR non codificanti.



La zona centrale è suddivisa in **esoni** (porzioni che contribuiscono alla sintesi della proteina) ed **introni**, che invece non portano informazioni.
Quando avviene la trascrizione (nel nucleo negli eucarioti), si ha l'aggiunta di guanina su 5' e adenina su 3'. In seguito, si eliminano gli introni ottenendo solo la parte codificante.

Rappresentazione grafica della sequenza trascritta



Esoni: Rettangoli

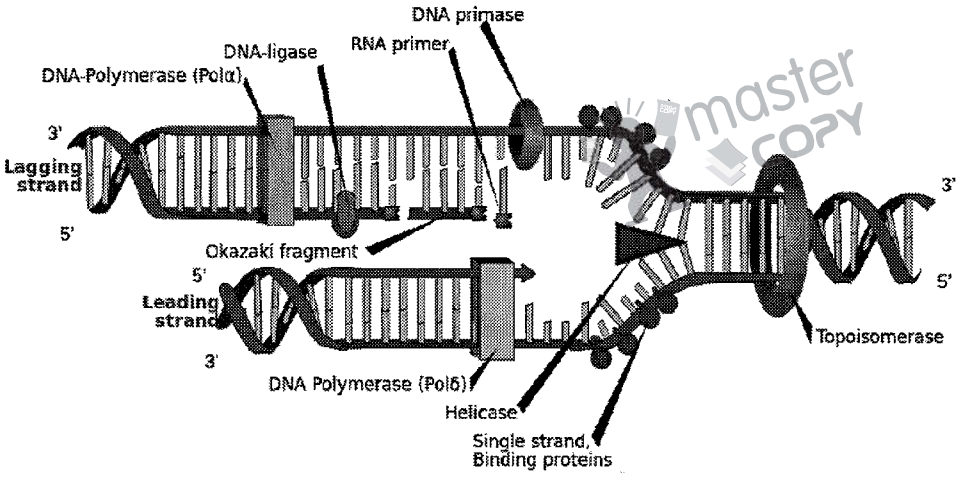
Sequenze codificanti (CDS): Rettangoli colorati

Sequenze non tradotte (UTR): Rettangoli non riempiti

Introni: Linee di connessione

Replicazione del DNA

La replicazione avviene aprendo la doppia elica, in seguito ogni filamento fa da stampo per la costruzione di quello corrispondente. Per aprire la doppia elica vengono utilizzati degli enzimi, e su ogni filamento va ad agire una polimerasi.
I due filamenti hanno versi di percorrenza opposti: su uno la polimerasi segue direttamente l'**elicasi**, nell'altro invece la polimerasi lavora nel verso opposto quindi lavora a step: ricrea una parte del filamento e si stacca per poi riprendere. Questo potrebbe portare all'insorgenza di buchi nella sequenza, chiamati **isole di Okazaki**.



Il meccanismo di replicazione del DNA è stato sfruttato in laboratorio per replicare una certa sequenza del genoma. Questa procedura è chiamata **PCR**.