

Le simmetrie del DNA n. 2

Piano Nazionale Lauree Scientifiche.
Progetto Math-en-Jeans.
Dipartimento di Matematica
Università di Padova.

Mario Puppi, Valentina Novello

18 gennaio 2019

4. Pattern e statistiche del DNA.

 **4.1 La funzione di autocorrelazione.** Fissiamo l'attenzione su un oligonucleotide ω cioè una parola formata con le lettere dell'alfabeto genetico $\mathcal{B} = \{A, C, T, G\}$. Fissato un *ritardo* t , consideriamo il pattern $C_\omega(t) = \omega \dots \omega$ che inizia e termina con la stessa parola ω e in mezzo una sequenza arbitraria di t lettere.

Ad esempio, $C_{ACT}(1) = ACT * ACT$, $C_{ACT}(2) = ACT * * ACT$

La legge di simmetria per la funzione di autocorrelazione diventa: $C_\omega(t) = C_{\hat{\omega}}(t)$

 **4.2 La distribuzione tempo-di-ricorrenza.** Dati un oligonucleotide ω e una distanza t , definiamo $R_\omega(t)$ come il tempo del primo ritorno ossia la distanza minima tra due occorrenze della parola ω .

Si può dimostrare, usando aritmetica e combinatoria elementari, che $R_\omega(t)$ può essere scritto come una somma di differenti frequenze $P(X)$

La legge di simmetria predice che $R_\omega(t) = R_{\hat{\omega}}(t)$