Approcci di Programmazione Lineare Intera al problema della ricombinazione del DNA nei ciliati

Antonio Vivace

Panoramica del lavoro

Programmazione Lineare Intera. Definizione, tecniche e algoritmi risolutivi. Relazione con la bioinformatica. Traduzione di idiomi logici in sistemi di disequazioni lineari.

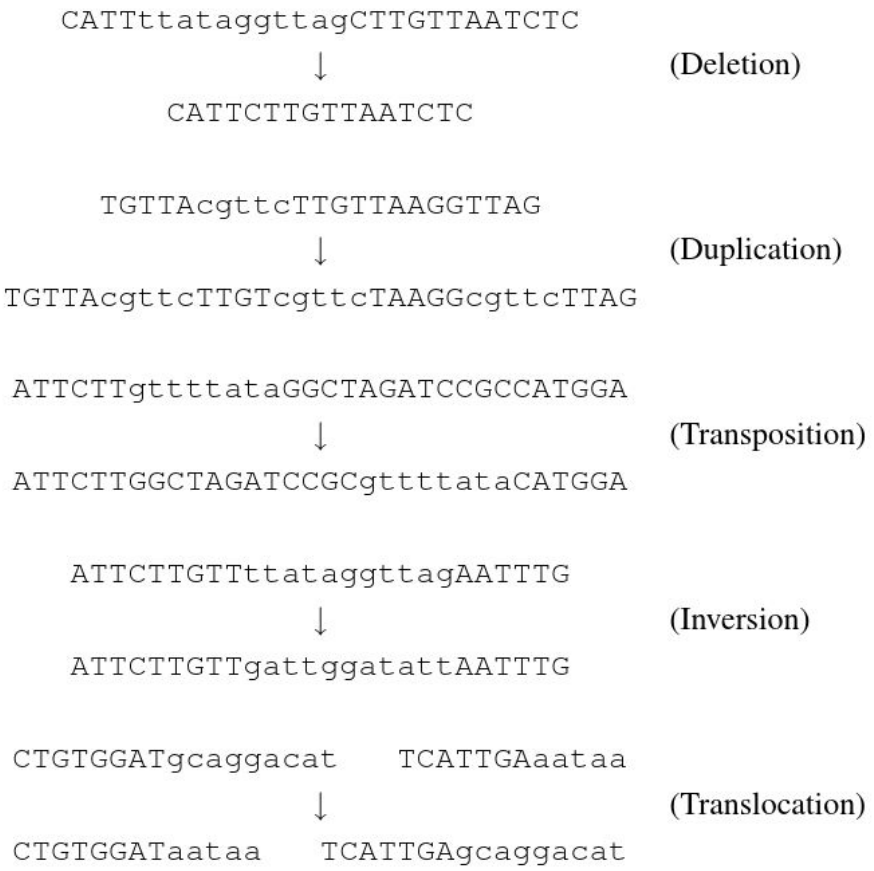
Il problema della ricombinazione del DNA e la sua manifestazione nei ciliati. Contesto biologico e motivazione. Formalizzazione e panoramica degli approcci esistenti.

Sperimentazione. Tentativi di formulazione del problema in termini di ILP. Software per la generazione procedurale di istanze ridotte. Proposta di un formato che rappresenti le mappe di riarrangiamento. Stesura del problema in termini di un un risolutore commerciale di problemi di programmazione lineare.

Eventi evolutivi

Computazione della *Distanza Evolutiva*

Confronti di interi genomi per evidenziare gli eventi evolutivi che li separano.



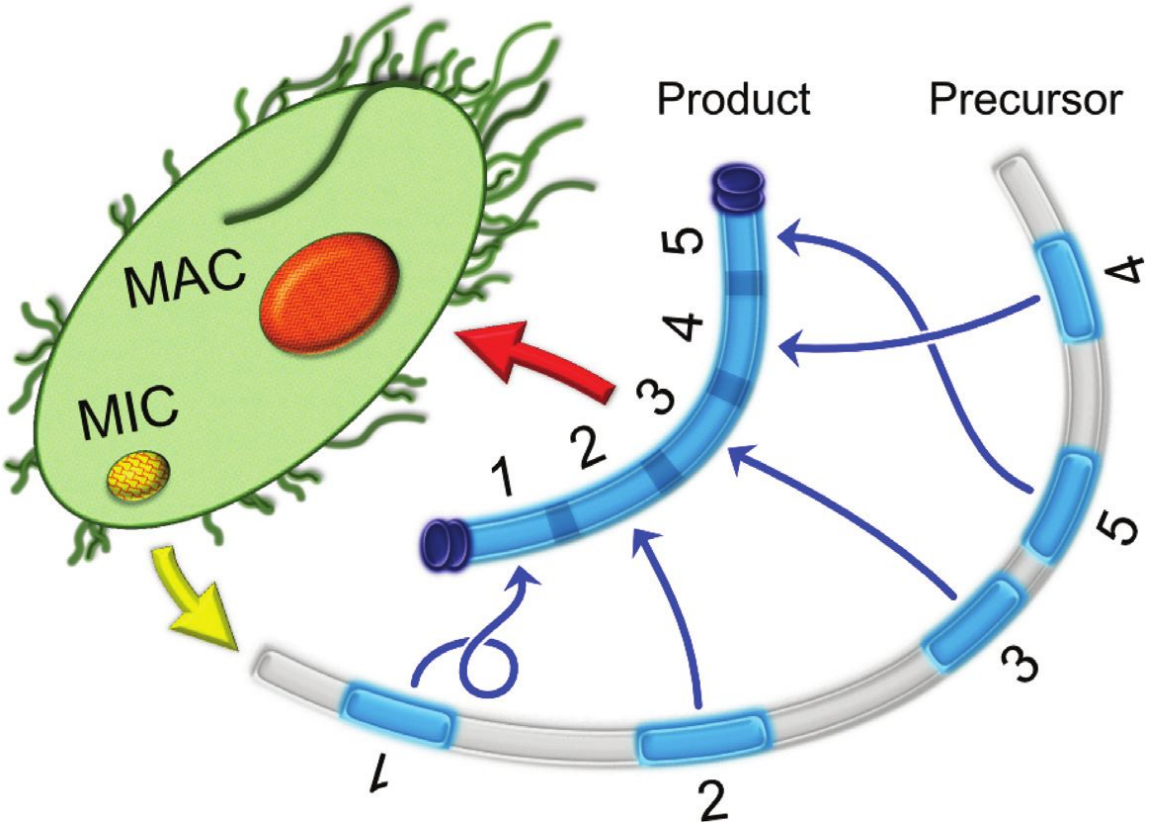
Riarrangiamento del DNA

Dato un insieme di **genomi** e un insieme di possibili **eventi** evolutivi (operazioni), trovare il più piccolo insieme di eventi che trasforma un genoma in un altro.

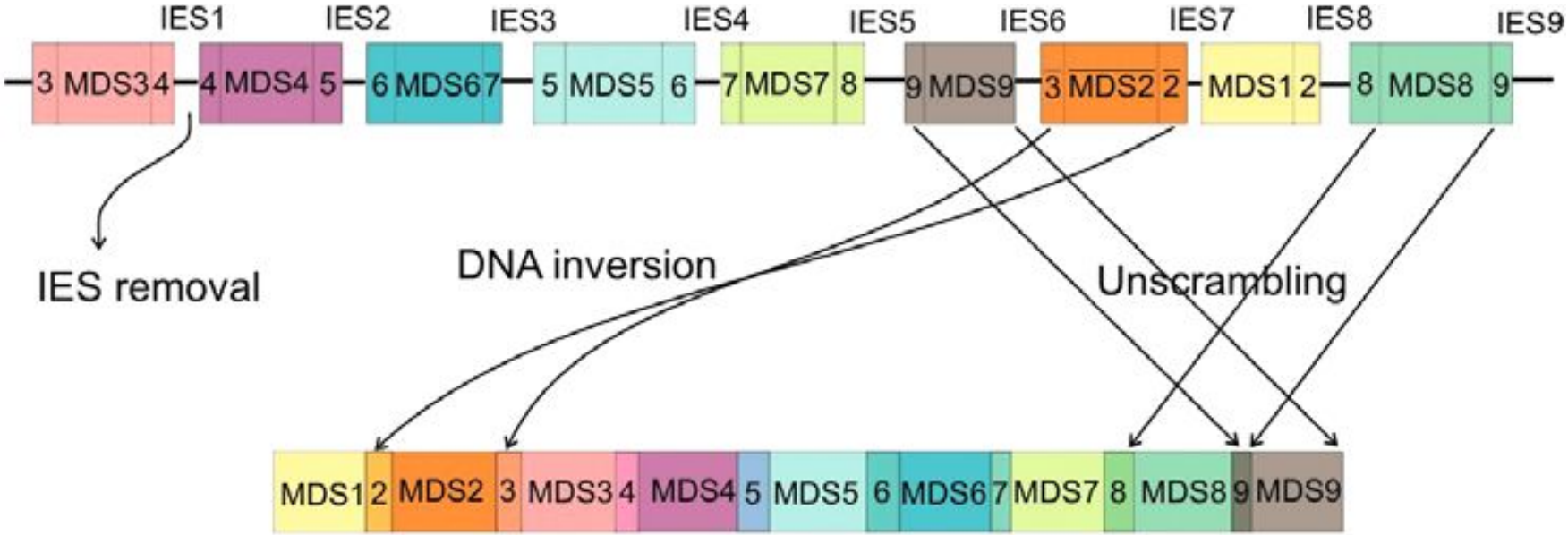
Diversità del problema: **cosa** rappresentano i genomi e **quali** sono gli **eventi possibili**.

Eventi rari?

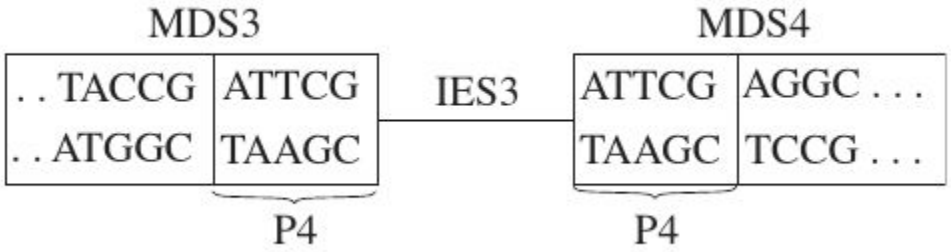
...nei ciliati



Eventi possibili



Sezioni di sovrapposizione



Programmazione Lineare Intera

I problemi di Programmazione Lineare Intera sono molto più *difficili* di quelli di Programmazione Lineare. Nessun algoritmo di soluzione generale è conosciuto ma ci sono dei risolutori (commerciali) veloci in pratica.

Strumento per descrivere, ri-formulare problemi (in bioinformatica)

Algoritmi Esatti, Euristici e Approssimativi

$$
\begin{aligned}
&\text{maximize} && \mathbf{c}^{\mathrm{T}}\mathbf{x} && \text{(cost function)} \\
&\text{subject to} && A\mathbf{x} \leq \mathbf{b} \\
&\text{and} && \mathbf{x} \geq \mathbf{0} \\
& && (\mathbf{x} \in \mathbb{Z}^n)
\end{aligned}
$$

Idiomi di logica come disequazioni lineari

If-Then	Only-If
$L \geq b \rightarrow z$	$z = 1 \text{ only if } L \geq b$
$L - (M \times z) \leq b - 1$	$L + m \times z \geq m + b$

Con ***L*** funzione lineare limitata superiormente da ***M***, inferiormente da ***s*** con ***m = s - b.***
b numero positivo intero.

Sperimentazione

La formulazione ILP proposta consiste in 12 variabili e 14 vincoli lineari.

Risolve istanze del problema da 50-100 caratteri, generate proceduralmente con uno script che ne simula le caratteristiche in modo più o meno accentuato.

Una fase preliminare, chiamata *preprocessing*, si occupa di popolare alcune variabili facendo pattern matching sull'istanza.

La **mappa di riarrangiamento** viene ricostruita a partire dalla soluzione del problema utilizzando gli assegnamenti alle variabili.

Alcuni idiomi della formulazione proposta

Variabili

$$MDS_{MACstart}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ starts at position } j \text{ in the MAC} \end{cases}$$

$$*Eq(i,j,h,l) = \begin{cases} 0 \\ 1, & \text{if } \texttt{MIC[i:j]} = \texttt{MAC[h:l]} \end{cases}$$

Objective Function: $min \sum_{i,j} MDS_{MACstart}(i,j)$

Vincoli

(2) $MDS_{MICstart}(i,a)+MDS_{MICend}(i,b)+MDS_{MACstart}(i,c)+MDS_{MACend}(i,d) \leq 4Eq(a,b,c,d) \quad \forall i,a,b,c,d$

(3b) $\sum_{j} MDS_{MACstart}(i,j) \leq 1 \quad \forall i$

Conclusioni

Approcci ibridi

Maggiori dettagli sul processo biologico

Panoramica dei metodi utilizzati

Costruzione di software che genera proceduralmente istanze con le volute caratteristiche e le pre-processa

Formato per l’annotazione di mappe di riarrangiamento