

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

Antonio Vivace

Panoramica del lavoro



Biologia (molecolare) computazionale, *Bioinformatica*. Domini applicativi, metodo, approcci e principali problemi affrontati.

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 risolutore commerciale di problemi di programmazione lineare.

Bioinformatica Biologia (molecolare) computazionale



Acidi nucleici ed evoluzione molecolare

Strutture e funzione delle proteine

TODO

Eventi evolutivi

Computazione della *Distanza Evolutiva*

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

CATTttataggtttagCTTGTTAATCTC



CATTCTTGTTAATCTC

(Deletion)

TGTTAcgttcTTGTTAAGGTTAG



TGTTAcgttcTTGTcgttcTAAGGcgttcTTAG

(Duplication)

ATTCTTggttttataGGCTAGATCCGCCATGGA



ATTCTTGGCTAGATCCGCgttttataCATGGA

(Transposition)

ATTCTTGTTttataggtttagAATTTG



ATTCTTGTTgattggatattAATTTG

(Inversion)

CTGTGGATgcaggacat TCATTGAaataa



CTGTGGATaataa TCATTGAgcaggacat

(Translocation)

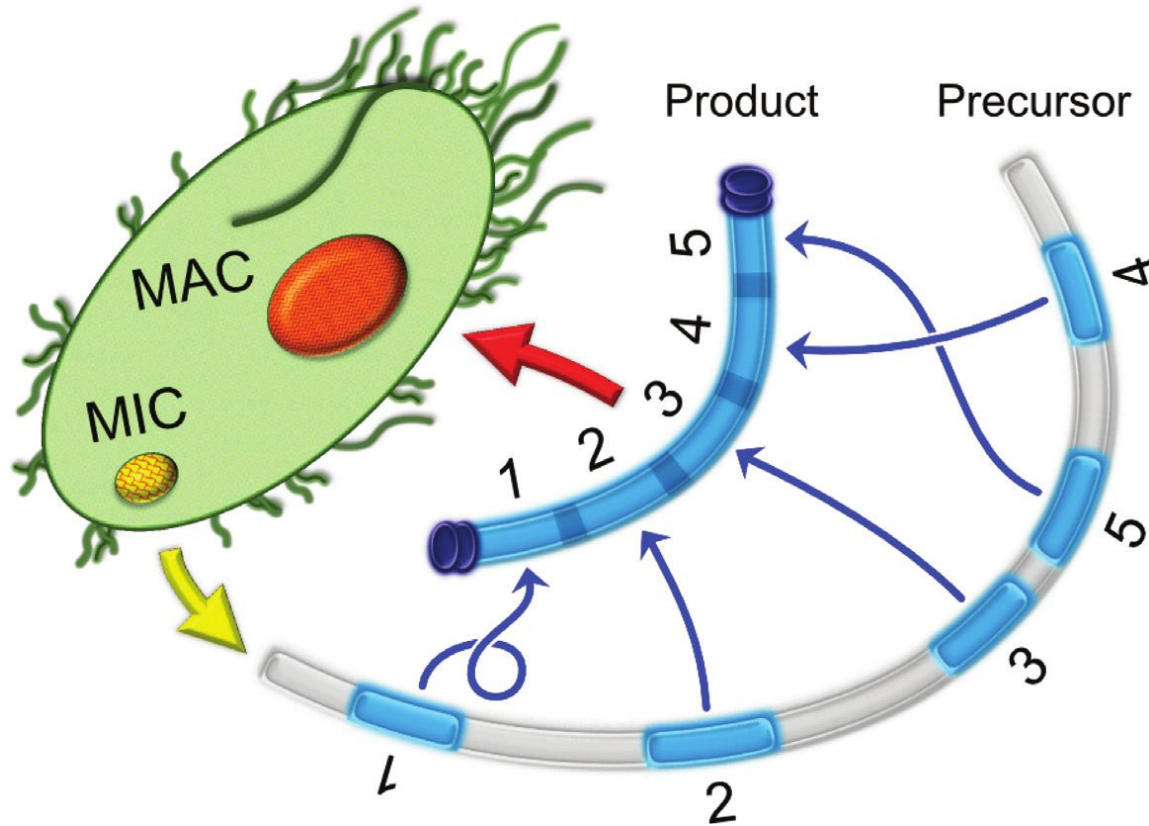
Riarrangiamento del DNA



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.

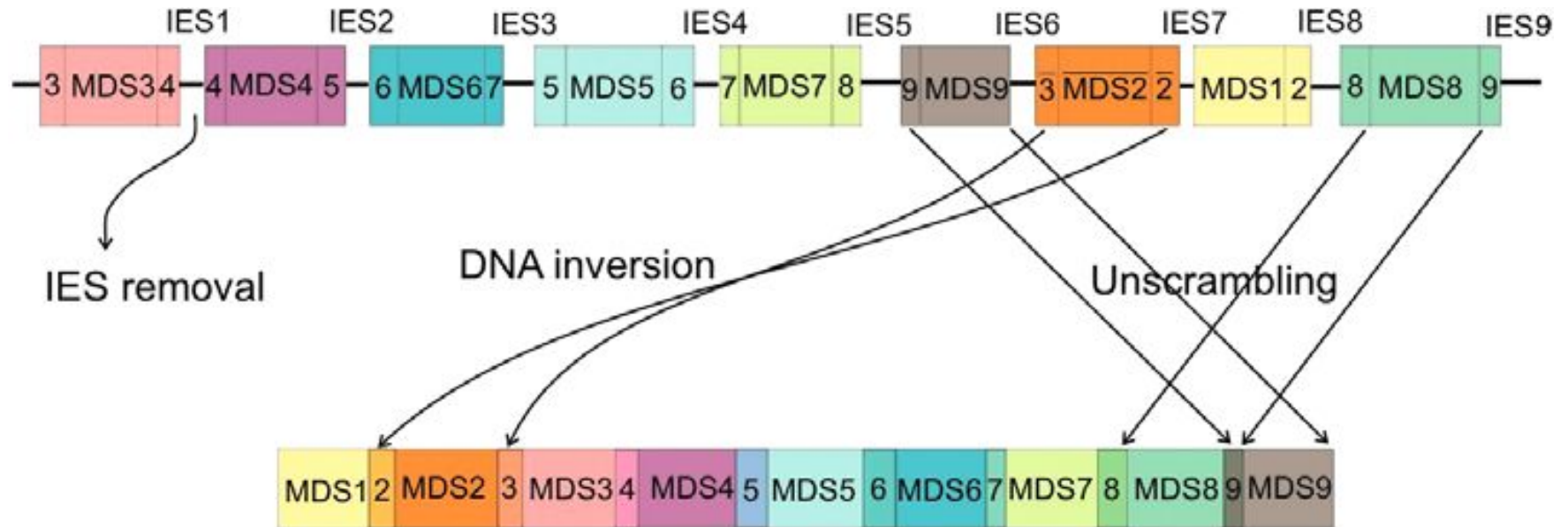
Diversità del problema: **cosa** rappresentano i genomi e **quali** sono gli **eventi** possibili.

Eventi rari?

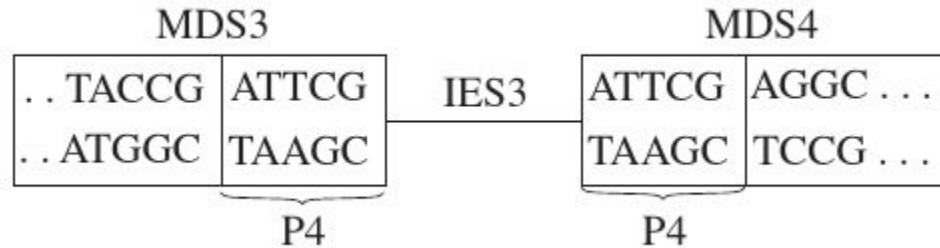


...nei ciliati

Eventi possibili



Sezioni di sovrapposizione



Programmazione Lineare Intera



NP HARD I problemi di Programmazione Lineare Intera sono molto più *difficili* di quelli di Programmazione Lineare. Nessun algoritmo di soluzione generale è conosciuto.

Strumento per descrivere, riformulare problemi in bioinformatica

Algoritmi Esatti, Euristici e Approssimativi

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

Idiomi di logica come disequazioni lineari



If-Then

$$L \geq b \rightarrow z$$

$$L - (M \times z) \leq b - 1$$

Only-If

$$z = 1 \text{ only if } L \geq b$$

$$L + m \times z \geq m + b$$

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

Alcuni idiomi della formulazione proposta

Variabili

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

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

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

Vincoli

$$(2) \quad \frac{MDS_{MICstart}(i, a) + MDS_{MICend}(i, b) + MDS_{MACstart}(i, c) + MDS_{MACend}(i, d)}{4Eq(a, b, c, d)} \leq 1 \quad \forall i, a, b, c, d$$

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

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