Università degli Studi di Milano-Bicocca

Identificazione efficiente di esoni

Relatore

Prof. Gianluca Della Vedova

Correlatore

Dott. Luca Denti

Candidato

Davide Cozzi

24 Luglio 2020

Introduzione e scaletta

Prerequisiti:

- concetti di biologia molecolare
- strumenti computazionali:
  - splicing graph, linearizzazione e bit vector
  - MEMs e MEMs graph

Introduzione e scaletta

Prerequisiti:

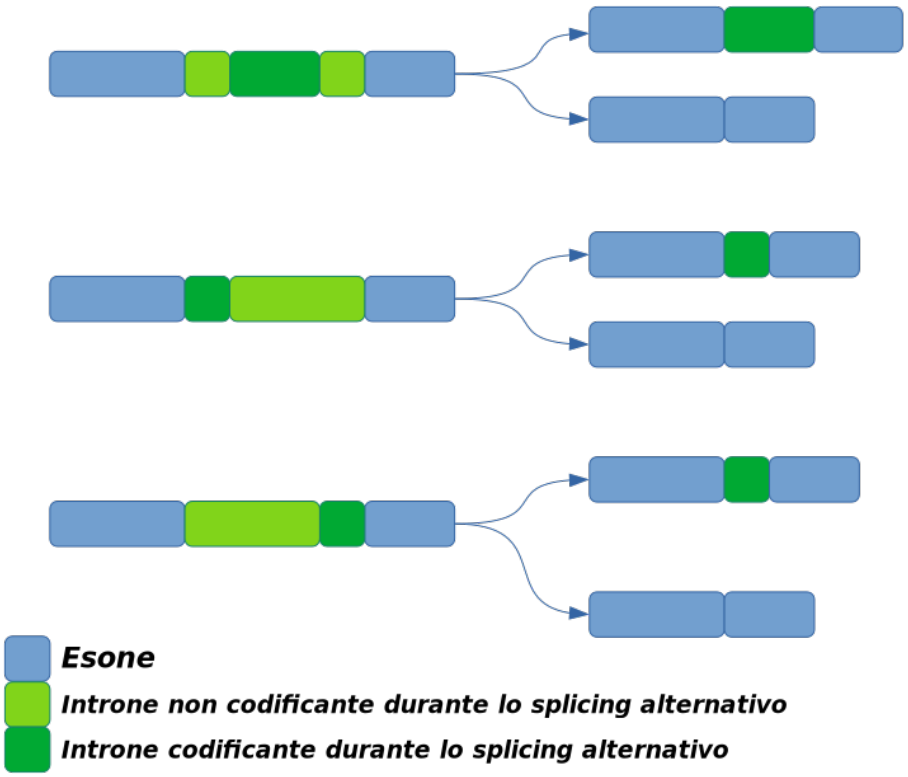
- concetti di biologia molecolare
- strumenti computazionali:
  - splicing graph, linearizzazione e bit vector
  - MEMs e MEMs graph

Innovazioni, riconoscimento di *novel exons* in ASGAL:

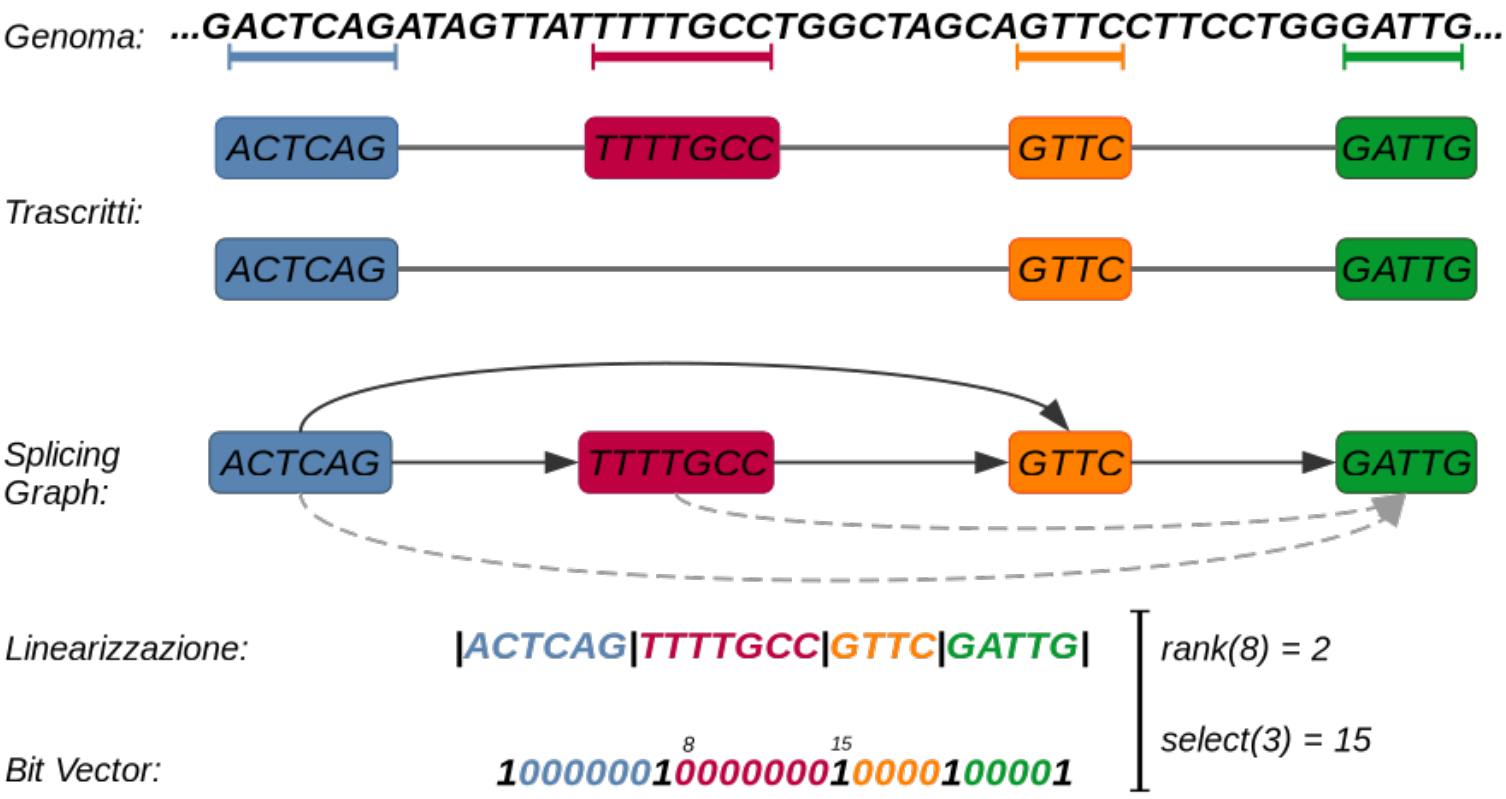
- riconoscimento degli introni
- estensione o ricostruzione del MEMs graph
- analisi dei risultati
- conclusioni e prospettive future

Prerequisiti: accenni di biologia molecolare

- *DNA e RNA*
- *sintesi proteica*
- *esoni e introni*
- *splicing alternativo*

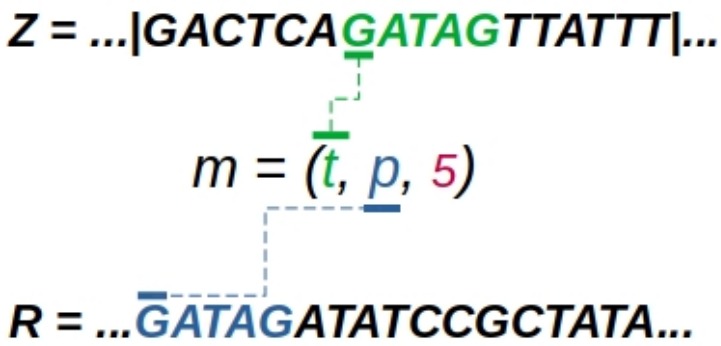


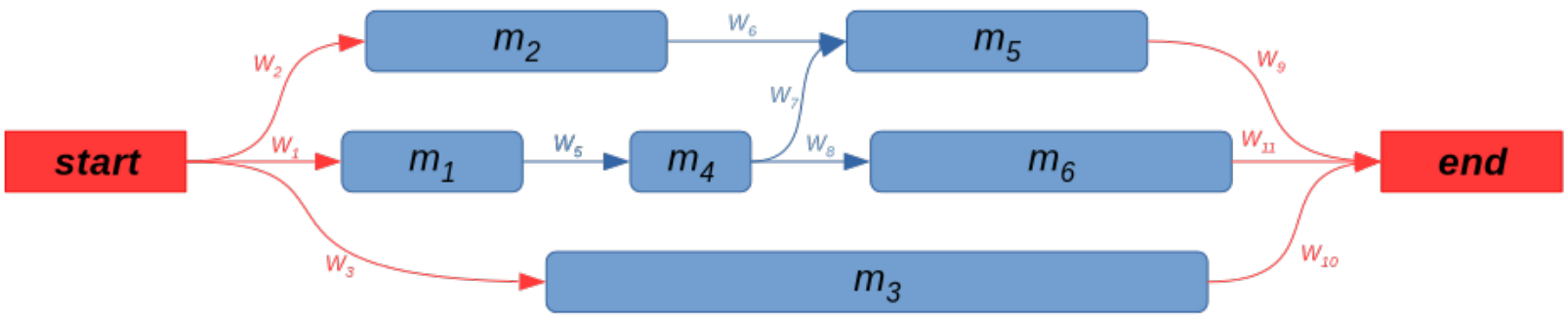
Prerequisiti: splicing graph, linearizzazione e bit vector



Prerequisiti: Maximal Exact Matches e MEMs graph

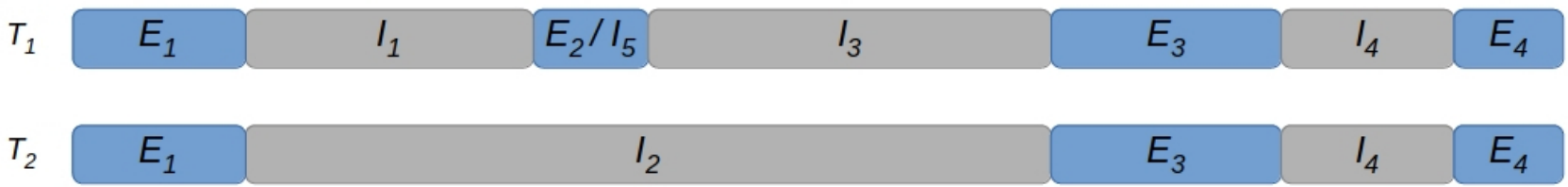
$m = (t, p, l)$





Innovazioni: riconoscimento degli introni

- riconoscimento degli introni a partire dall'annotazione
- costruzione della linearizzazione degli introni
- costruzione della mappa degli introni

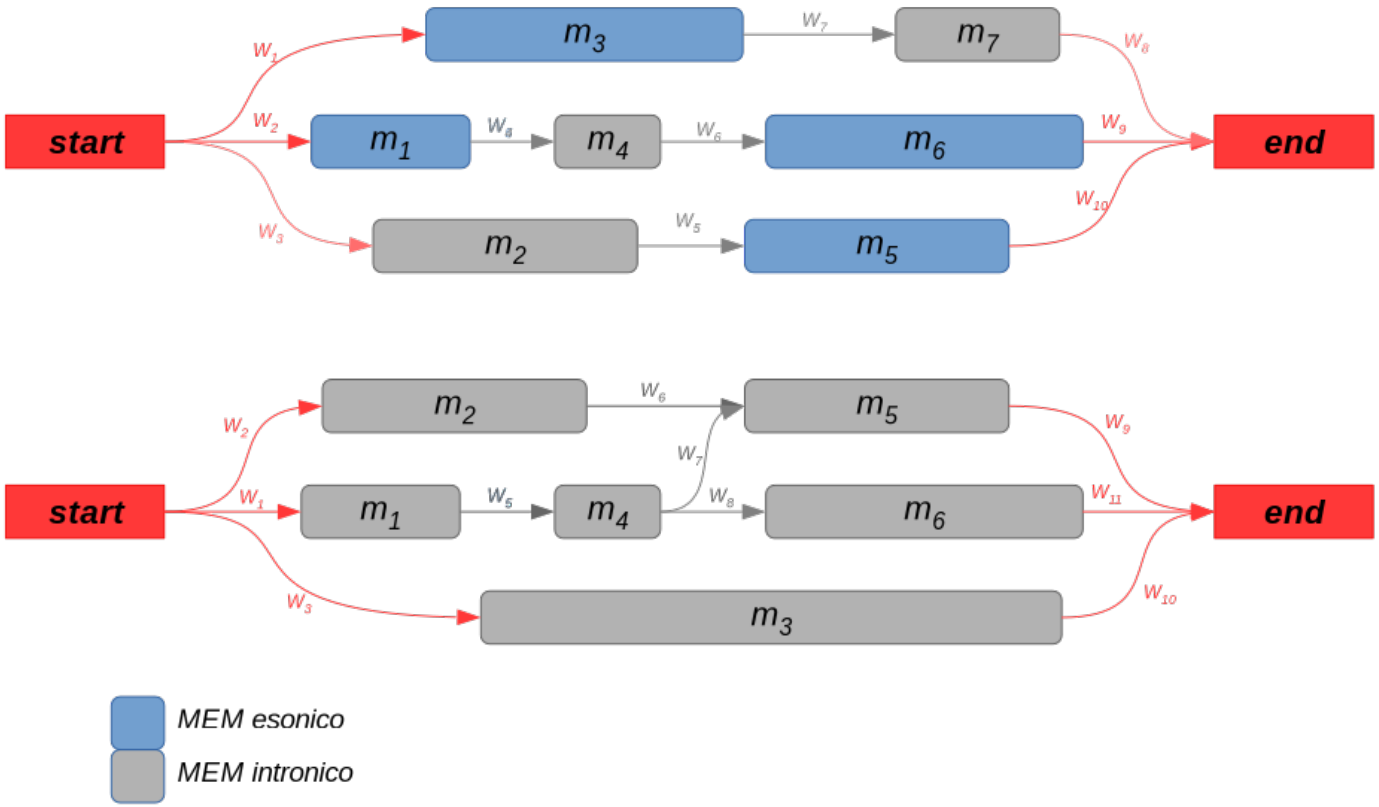


Mappa degli introni associata ai due trascritti

- $(E_1, E_2) = [I_1]$
- $(E_1, E_3) = [I_1, I_2, I_3, I_5]$
- $(E_2, E_3) = [I_3]$
- $(E_3, E_4) = [I_4]$

Innovazioni: estensione del MEMs graph

$$m = (t,\ p,\ l,\ \{0,\ 1\})$$



Innovazioni: sperimentazione e analisi dei risultati

- download del genoma e dell'annotazione
- manipolazione dell'annotazione
- simulazione dell'RNA-Seq sample (25000 reads lunghe 100)
- calcolo dei **MEMs**
- produzione e analisi dei risultati

- automatizzazione con Snakemake
- semplice modifica dei parametri
- parallelismo

cromosoma	error rate	tipo	match	$\leq$10%	$>$10%	mismatch	tempo (s)	memoria (Mb)
9	1%	intron	24747	194	23	36	9.6	59.1
		noIntron	22993	64	6	1937	4.6	57.2
9	2%	intron	24748	194	23	35	9.6	59.2
		noIntron	22993	64	6	1937	4.6	57.9
9	10%	intron	24748	193	23	36	10.0	49.7
		noIntron	22991	64	6	1939	4.6	49.5

Conclusioni e prospettive future

Considerazioni sui risultati

- ✓ buona qualità dei risultati
- ✗ performances (tempi medi di esecuzione) non ottimali

Conclusioni e prospettive future

Considerazioni sui risultati

✓ buona qualità dei risultati

✗ performances (tempi medi di esecuzione) non ottimali

Prospettive future

- ottimizzazione dello studio degli introni
- perfezionamento del **MEMs graph**
- parallelizzazione dello studio delle reads

Ringraziamenti

Grazie per l'attenzione

Relatore

Prof. Gianluca Della Vedova

Correlatore

Dott. Luca Denti

Candidato

Davide Cozzi

24 Luglio 2020