



Università degli studi Milano Bicocca
Scuola di Scienze

Dipartimento di Informatica , Sistemistica e Comunicazione
Corso di Laurea in Informatica

Sviluppo di una libreria Python per la gestione di grafi del pangenoma

Relazione della prova finale di:

Pizzo Davide

Matricola: 869254

Relatore: Prof. Gianluca Della Vedova

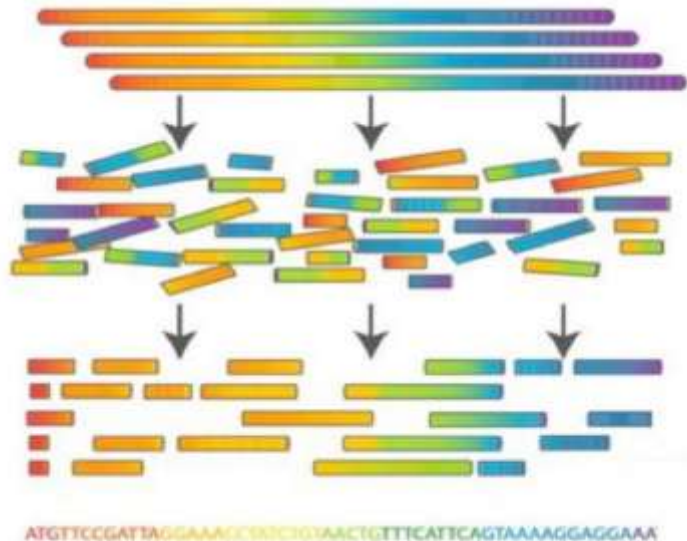
Co-relatore: Dott. Simone Ciccolella

Anno Accademico 2022/2023

Il contesto: assemblaggio di genomi e grafi di assemblaggio

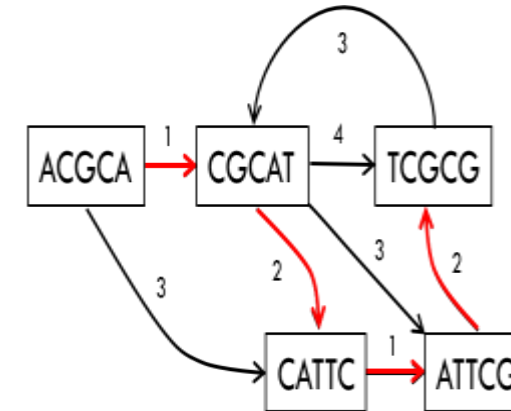
Assemblaggio genomico

- Processo di ricostruzione di un genoma, partendo da frammenti di DNA
- Comprensione della struttura genomica



Grafi di assemblaggio

- Strumento chiave nell'assemblaggio di genomi
- Rappresentazione grafica delle relazioni tra frammenti di DNA



ACGCATTCGCG

Obiettivi

❑ Estendere la libreria pygfa

Cosa è stato fatto

- Implementate diverse funzioni per la gestione e interrogazione del grafo
- Funzioni sottoposte a una fase di testing

❑ Riduzione uso memoria

Cosa è stato fatto

- Modifica gestione paths

❑ Intraprendere lo sviluppo del tool GFATK

Cosa è stato fatto

- Sviluppato il design e il funzionamento
- Implementato un primo comando

Il formato GFA

- ❑ Ogni riga rappresenta un elemento del grafo
- ❑ Il primo campo di ogni riga definisce il tipo di essa

S	17	CTTACG	LN:i:6
---	----	--------	--------

L	1	+	106	-	7M
---	---	---	-----	---	----

P	14	11+,12-,13+	4M,5M
---	----	-------------	-------

Risparmio memoria

Sottografo GFA



Lista annidata

- ❑ Gestione dei cammini

- ❑ Modifica metodo 'from_string'

- ❑ Maggiore semplicità lato
utente/sviluppatore

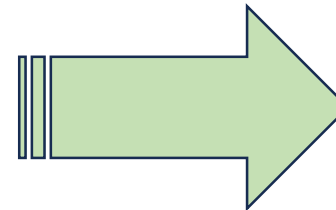
Gestione stringhe

Scopo: Ottenere la concatenazione delle sequenze di un cammino, tenendo conto dell'orientamento

❑ Input : ID path

❑ Output : Sequenza concatenata

S	1	ACCTT	
S	2	AGCTAA	
S	3	ATGA	
S	4	ATTGGCC	
S	5	TACG	
P	10	1+, 2-, 5+	2M, 1M



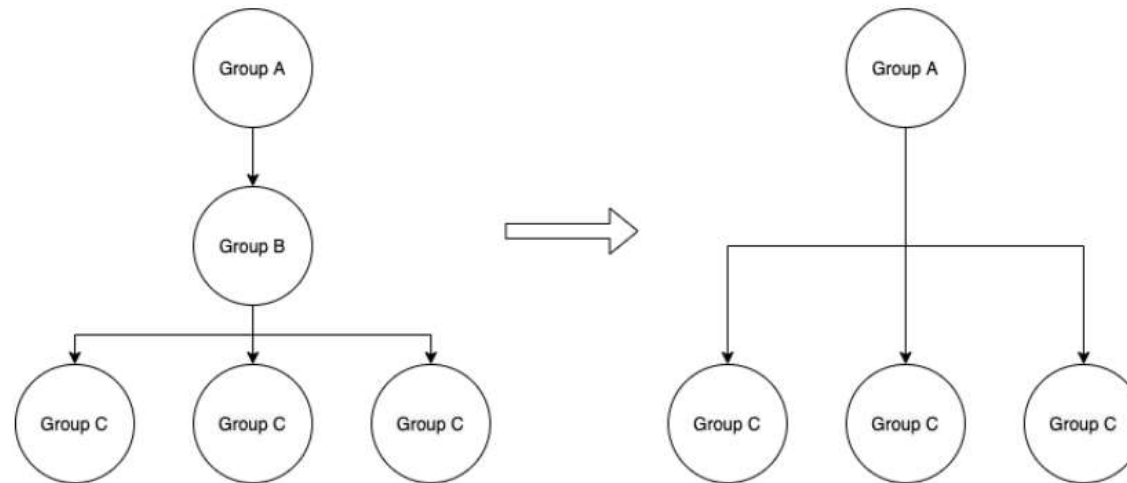
ACCTTTTAGCTTAGC

Gestione grafo

Scopo: Eliminare un nodo e creare gli archi di collegamento tra predecessori e successori, dati in input l'ID del nodo e la modalità

Modalità 1

Modalità 2



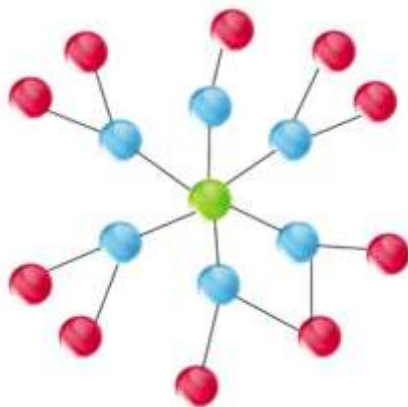
Archi di collegamento tra tutti i predecessori e successori

Arco di collegamento in base ai cammini

Interrogazione grafo

Scopo: Interrogare il grafo e ottenere un sottografo che risponde a tale richiesta.

- ❑ Sottografo costruito partendo da un nodo e il suo intorno
- ❑ Sottografo costruito tra due nodi



- Nodo passato in input
- Intorno lungo 1 (corrispondente al vicinato)
- Intorno lungo 2

User interface

- ❑ Fornire un'esperienza d'uso intuitiva ed accessibile
- ❑ Progettato il tool GFATK
- ❑ Python gfatk <command> <args> <outputfile>
- ❑ Implementato un primo comando : dump
- ❑ Semplicità di utilizzo per gli utenti

Conclusioni e sviluppi futuri

Possibili sviluppi futuri

- ☐ Estendere pygfa con nuove funzionalità (es. Rimozione nodi isolati,aggiornamento sequenze...)
- ☐ Sviluppo gfatk con nuovi comandi (es. Convertire una sequenza in formato FASTA,...)
- ☐ Possibili modifiche a funzioni già implementate nella libreria (es. Modifica `get_Subgraph_between_nodes`)

Conclusioni

- ☐ La libreria è stata arricchita da nuove funzionalità
- ☐ Funzioni implementate correttamente
- ☐ Funzioni testate accuratamente

GRAZIE PER L'ATTENZIONE
