



# Quarta prova d'esonero

Mercoledì 27 Maggio 2020

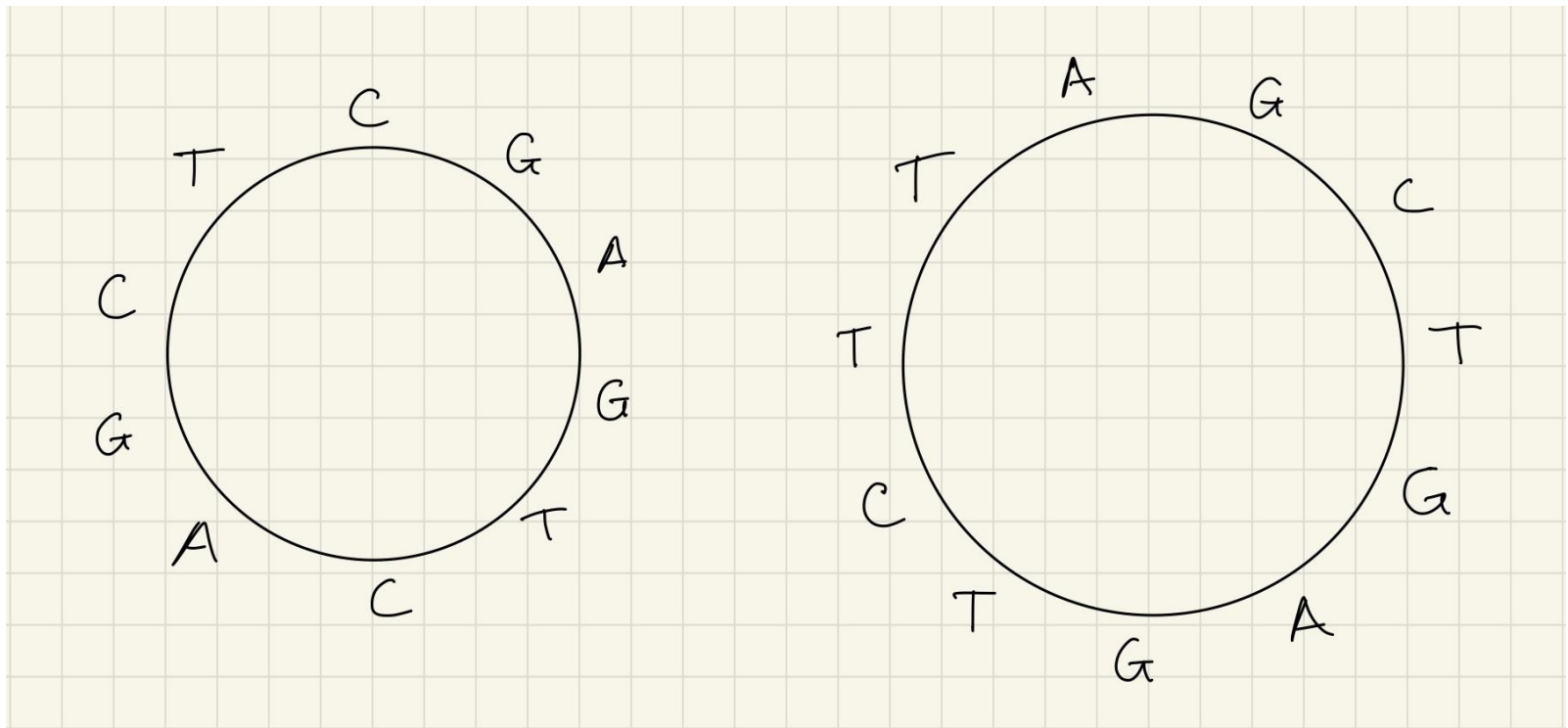
# Quarto esercizio di esonero

(da consegnare entro il 4 Giugno 2020 via mail)

Alcune sequenze di DNA esistono in forme circolari (per esempio nei virus) come nella figura, che mostra una sequenza circolare CGAGTCAGCT, cioè l'ultimo simbolo T in CGAGTCAGCT è collegato al primo simbolo C. Solitamente una sequenza circolare viene letta in senso orario. Un'azienda ha deciso di memorizzare le sequenze circolari di DNA come sequenze lineari. Tuttavia, ci possono essere molte sequenze lineari che sono ottenute da una sequenza circolare tagliando un punto qualsiasi della sequenza circolare. Quindi, l'azienda ha deciso di memorizzare la sequenza lineare che è lessicograficamente più piccola tra tutte le sequenze lineari che possono essere ottenute da una sequenza circolare. L'azienda inoltre, date due sequenze circolari, al fine di cogliere similarità negli organismi esaminati, vuole sapere qual è la più lunga sottostringa che le due sequenze circolari hanno in comune. Il vostro compito è quello di implementare gli algoritmi necessari all'azienda per i seguenti obiettivi:

1. individuare la sequenza lessicograficamente più piccola da una data sequenza circolare. Per l'esempio nella figura, la sequenza lessicograficamente più piccola è AGCTCGAGTC. Se ci sono due o più sequenze lineari che sono lessicograficamente più piccole, se ne sceglie una qualsiasi
2. date due sequenze circolari individuare la lunghezza della più lunga sottostringa che hanno in comune. Per l'esempio della figura tale lunghezza è 5
3. visualizzare una sottostringa la cui lunghezza è definita nel punto 2. Per l'esempio in figura tale sottostringa è GAGTC

# Figura



# Input e Output

L'input consiste in un 2 sequenze circolari di DNA, rappresentate in modo lineare. Ogni sequenza è preceduta da una riga che inizia con il simbolo > seguita dal nome della sequenza, per esempio >seq1

Ogni sequenza ha almeno 2 caratteri e può contenere solo i simboli A, C, G, T.

Il programma deve produrre in output:

1. Per ognuna delle due sequenza, la sequenza lineare lessicograficamente più piccola.
2. La lunghezza della più lunga sottostringa comune contenuta nelle due sequenze circolari
3. La più lunga sottostringa comune contenuta nelle due sequenze circolari.

# Esempio

## INPUT

>seq1

CGAGTCAGCT

>seq2

TTAGCTGAGTC

## OUTPUT

LexMin

>seq1

AGCTCGAGTC

>seq2

AGCTGAGTCTT

LCS

5

GAGTC

# Modalità di consegna

- + Spedire la soluzione entro e non oltre il 04/06/2020 via mail a [marinella.sciortino@unipa.it](mailto:marinella.sciortino@unipa.it) con oggetto:  
*Quarta prova esonero Lab Algoritmi – Nome Cognome*  
includendo in allegato in formato zip tutte le classi ( i file .java) necessarie alla compilazione e eventuali file di esempio prodotti. Si precisi qual è la complessità di tempo e di spazio dell'algoritmo proposto, descrivendo con opportuni commenti la strategia algoritmica proposta e le eventuali strutture dati ausiliarie utilizzate. Ogni metodo o attributo inserito nella classe deve essere adeguatamente commentato.