

تمرین سری پنجم بیوانفورماتیک

نکته مهم : تمرین پنجم در سایت کوئرا قرار داده شده است و شما می‌توانند از طریق [این لینک](#) در کلاس مربوطه در کوئرا با کلید Bio1400 ثبت نام نمایید. تحویل پاسخ‌های تشریحی نیز از طریق همین صفحه‌ی کوئرا انجام خواهد شد.

سوال (۱)

الف) برای مجموعه رشته‌های پروتئینی زیر، یک موتیف به صورت عبارت منظم بنویسید.

AYGTTSKK

AYPTTSIK

AVHTTSIK

AYMTTSIK

AVZTTTSIK

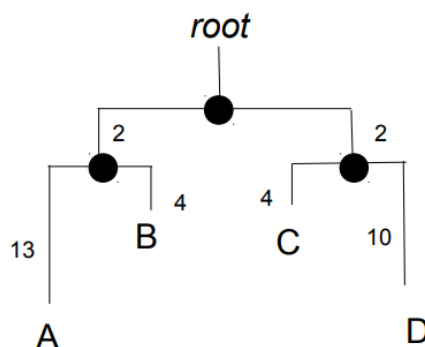
ب) عبارت منظم مقابل با کدام یک از رشته‌های زیر exact match می‌شود. برای هر رشته دلیل بیاورید.

$M - [TG] - X - \{M\} - A(2) - P - [YPC]$

- MMTGAAPP
- MTTTAAPC
- MGTMAAPP
- MTGAAPPY

سوال (۲)

به درخت حقیقی زیر که فاصله ۴ گونه از یکدیگر را نشان می‌دهد، دقت کنید. در این درخت فواصل هر دو گونه، از مجموعه فواصل شاخه‌های بین به دست می‌آید. به طور مثال فاصله گونه A تا C برابر با $4 + 2 + 2 + 13 = 21$ می‌باشد.



الف) ابتدا فواصل بین تمامی گونه‌ها را استخراج کرده و ماتریس فاصله را رسم کنید.

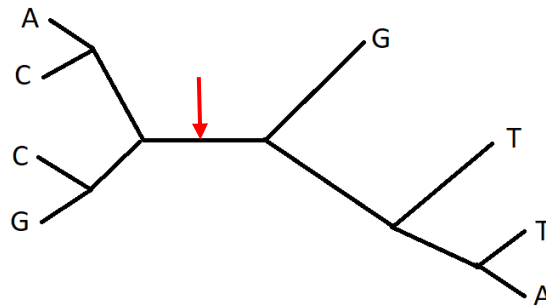
ب) با استفاده از ماتریس فاصله بدست آمده در قسمت (الف)، به روش UPGMA درخت را رسم کنید.

ج) با استفاده از ماتریس فاصله بدست آمده در قسمت (الف)، به روش NJ درخت را رسم کنید.

د) درخت بدست آمده در قسمت (ب) و (ج) را با درخت اصلی مقایسه کنید و در صورت وجود تفاوت، دلیل را بیان کنید.

سوال ۳)

فرض کنید درخت زیر ریشه دار است. با در نظر گرفتن ریشه در نقطه قرمز، حداقل تعداد جهش‌ها در درخت زیر را پیدا کنید و بررسی کنید که در صورت تغییر محل ریشه درخت آیا تغییری در تعداد جهش‌ها اعمال می‌شود یا خیر.



سوال ۴)

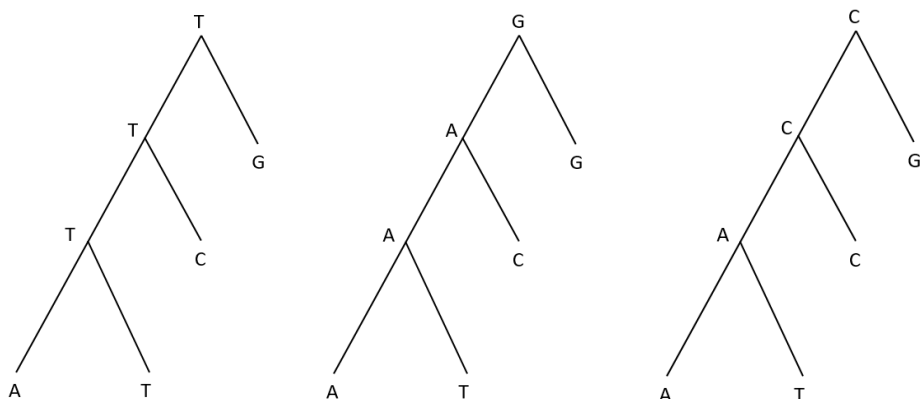
برای سوالات پارسیمونی بزرگ، دو روش exhaustive و branch-and-bound را به طور خلاصه شرح دهید و مقایسه کنید. همچنین اطمینان برای پیدا کردن اپتیمم سراسری را در این دو الگوریتم بررسی کنید.

سوال ۵)

احتمال رخداد درخت‌های زیر را طبق ماتریس احتمال جهش داده شده پیدا کنید.

From/ To	A	C	G	T
A	0.55	0.2	0.15	0.1
C	0.05	0.7	0.15	0.1
G	0.15	0.05	0.6	0.2
T	0.25	0.05	0.1	0.6

در نظر داشته باشید که $P(A) = P(C) = P(G) = P(T) = 0.25$ و برای مثال احتمال $P(A \rightarrow C) = 0.2$



نکات تحویل تمرین

- (۱) تمرینات را به صورت انفرادی انجام دهید.
- (۲) پاسخ های خود را با کیفیت مناسب و خوانا اسکن کرده و یا تایپ شده به صورت یک فایل درآورید و با نام "HW5_studentNumber_FirstnameLastname.pdf" در سایت بارگذاری نمایید.
- (۳) در صورت پیداشدن هرگونه کپی چه در سوالهای تشریحی و چه برنامه نویسی نمره ی هر دو نفر ۱۰۰- در نظر گرفته خواهد شد. این نکته را جدی بگیرید تا بعداً به مشکل نخورید.
- (۴) اشکالات خود را برای تمرین تشریحی می توانید از طریق ایمیل زیر بپرسید.
aref78.m@gmail.com
- (۵) مهلت ارسال این تمرین تا ۱۴۰۱/۴/۴ است و پس از این تاریخ حداکثر تا دو روز به ازای هر ساعت تاخیر ۱ درصد جریمه در نظر گرفته خواهد شد. بعد از این مدت پاسخ سوالات در کانال درس ارسال شده و دیگر نمره ای به تمرین تعلق نخواهد گرفت.