

نموذج A:

السؤال الأول: (٦ درجات)

يفرض لدينا الجينوم المرجعي التالي ATTGACATTAGAT

- ١- باستخدام القوائم المرتبة (Order List) ، قم بإنشاء (k-mer Index) للجينوم المرجعي السابق ومن أجل (k=3).
- ٢- باستخدام (k-mer Index) الذي تم بناؤه في الخطوة السابقة، قم بالاستعلام عن القراءات التالية لإيجاد الـ (index hits) الموافقة ومن ثم التحقق فيما إذا كانت هناك مطابقة تامة (exact match) أم لا.

P= GACATT , P1=TTAGTA

السؤال الثاني: (٤ درجات)

باستخدام لغة البرمجة Python:

- ١- قم بكتابة تابع يطبع النسبة المئوية لظهور كل Base (A, C, G, or T) من جينوم ما. على سبيل المثال، من أجل الجينوم التالي AACCGGTT: يجب أن يكون الخرج 25% لكل قراءة.
- ٢- قم بكتابة تابع يقوم بتحويل قيمة جودة القراءة الناتجة عن تقنية 2nd generation sequencer ما للشكل الذي يتم تخزينها به ضمن ملف FASTQ.

نموذج B:

السؤال الأول: (٦ درجات)

TAACTGTAATCTA

بفرض لدينا الجينوم المرجعي التالي

١. باستخدام القوائم المرتبة (Order List) ، قم بإنشاء (k-mer Index) للجينوم المرجعي السابق ومن أجل (k=3).
٢. باستخدام (k-mer Index) الذي تم بناؤه في الخطوة السابقة، قم بالاستعلام عن القراءات التالية لإيجاد الـ index (hits) الموافقة ومن ثم التحقق فيما إذا كانت هناك مطابقة تامة (exact match) أم لا.

P= CTGTAA , P1=AATCAT

السؤال الثاني: (٤ درجات)

باستخدام لغة البرمجة Python:

- ١- قم بكتابة تابع يعيد آخر موقع ظهور لـ pattern ما p ضمن جينوم مرجعي T، على سبيل المثال من أجل p=ACG و T= ACGTTGCACGTTGCACG يجب أن يكون الناتج هو الموقع 14.
- ٢- قم بكتابة تابع يقوم بتحويل قيمة جودة القراءة الناتجة عن تقنية 2nd generation sequencer ما للشكل الذي يتم تخزينها به ضمن ملف FASTQ.

وظيفة (Homework):

اكتب ملف Python يقوم بتنفيذ التالي:

- ليكن لديك سلسلة ما من المحارف ليكن 100 محرف مثلاً من مجموعة المحارف ('A', 'T', 'G', 'C'), والمطلوب تحديد فيما إذا كانت سلسلة ما 'ACTG' موجودة ضمن هذه السلسلة أم لا احتفظ بقائمة تحوي كل المواقع التي تظهر فيها السلسلة.
- ابحث من خلال توثيق لغة Python عن طريقة أو دالة تقوم بهذا العمل دون الحاجة لكتابة الدالة السابقة.

وظيفة Homework:

باستخدام لغة البرمجة Python قم بكتابة واختبار تابعين يقومان بالإجراءات التالية:

- ليكن لدينا سلسلتين DNA والمطلوب كتابة تابع يوجد أطول سلسلة مشتركة بينهما، فمثلاً من أجل ACCAGCGT و ACCAGTCG يكون الناتج ACCAG.
- باستخدام القواميس قم بكتابة تابع يعطي النسخة المقابلة لسلسلة DNA، فمثلاً من أجل السلسلة ACCGTCG يكون الخرج CGACGGT.
- قم بتحميل ملف FASTQ من اختيارك و قم بإعادة تطبيق القسم الموجود من فقرة FASTQ موضحاً النتائج التي تحصل عليها.

وظيفة Homework:

باستخدام لغة البرمجة Python وبالاعتماد على الأكواد السابقة قم بجلب ملف FASTA وملف قراءات FASTQ من اختيارك و قم بتطبيق الخوارزمية السابقة معللاً النتائج التي حصلت عليها.

وظيفة Homework:

- 1- أحد التعديلات المقترحة على خوارزمية الفهرسة المدروسة هو الاحتفاظ بسلاسل DNA الجزئية بطول k والتي تبدأ بموقع زوجي فقط بهدف تحقيق فهرس أصغر مما يؤمن بحثاً بشكل أسرع، المطلوب قم بكتابة الفهرس للسلسلة المذكورة وفق القيود السابقة. واختبر وجود السلسلة الجزئية التالية (GCGTGCTT) ضمن القراءة (CGTGCGTGCTT). كيف يمكنك إيجاد السلسلة في حالة كانت تبدأ بموقع فردي ضمن الجينوم؟؟؟
- 2- برأيك أي البنى المدروسة أفضل؟ قارن بين الطريقتين من حيث زمن تنفيذ كل منهما مع تعليل إجابتك.