

תרגיל בית מספר 2

חומר הלימוד לתרגיל: הקוד הגנטי, מוטציות, עימוד זוג רצפים, genbank

חלק 1 - הקוד הגנטי ומוטציות פוטנציאליות:

בקובץ בשם mutations.py כתבו פונקציה בשם

```
count_mutation_by_type(position, type)
```

הפונקציה עובדת על הקוד הגנטי (הגדירו אותו כמשתנה גלובלי בקובץ הקוד). הפונקציה מקבלת מספר שלם (position) המייצג עמדה בקודון (כלומר 1, 2 או 3), ומחזרת המייצגת סוג מוטציה (הערכים האפשריים הם "nonsense" או "synonymous"). הפונקציה מחשבת עבור הקודונים שאינם stop codons, כמה מוטציות אפשריות קיימות עבור העמדה המבוקשת, עבור סוג המוטציה המבוקשת. כך למשל, עבור הקריאה לפונקציה:

```
count_muation_by_type(position=1, type='nonsense')
```

תחזיר 9, אילו הקריאה לפונקציה:

```
count_muation_by_type(position=2, type='nonsense')
```

תחזיר 7, ואילו הקריאה לפונקציה:

```
count_muation_by_type(position=1, type='synonymous')
```

תחזיר 8 (במקרה הזה נספור את המוטציות בשני הכיוונים, בין שני הקודונים אשר מקודדים לאותה חומצת אמינו).

הערות:

- עבדו עם הקוד הגנטי שמופיע ב**קישור הזה**.
- ניתן להניח שהפרמטרים שיועברו לפונקציה יהיו חוקיים.

חלק 2 – עימוד רצפים:

בקובץ בשם alignment.py ממשו את המשימה הבאה:

1. נתונים שלושה רצפים של מקטעים גנטיים אשר נלקחו משלושה מינים שונים. קבעו מי הם שני המינים הקרובים יותר, על סמך דמיון בין הרצפים (% identity). כלומר, בצעו global pairwise sequence alignment וספרו התאמות. השתמשו בפרמטרים הדיפולטיביים של פונקציית העימוד globalxx.
ענו על השאלה עבור 2 קבוצות רצפים:
א. רצפי דנ"א מבבון, חתול וקוף-מקק
ב. רצפי דנ"א מיטוכונדריאלי מאדם, שימפנזה ואורנגאוטן

הסבירו את התוצאות בקובץ ex2.docx. האם התוצאות הגיוניות ומתאימות לציפיות?

הערות:

- מאחר שייתכן יותר מעימוד אחד אפשרי, בחרו את העימוד הראשון שמתקבל.
- הרצפים עבור סעיפים א' נמצאים בקובץ ה fasta המצורף.
- שליפת הרצפים עבור סעיף ב' תעשה בעזרת genbank. השתמשו ב- accession numbers הבאים עבור שליפת הרצפים:
"AF451972", "AF176731", "X90314"
אין לייצר קבצים נוספים בסעיף זה, כלומר שלפו ישירות את הרצף, ללא שמירת קובץ ביניים.

2. בהמשך לסעיף 1, עבור כל זוג רצפים, אפיינו את המוטציות השונות שחלו בהסתמך על העימוד המתקבל (השתמשו באותו עימוד שבו השתמשתם בסעיף 1: צרו מילון אשר סופר את סוגי העמדות השונות:

- Matches – מספר התאמות
- Mismatches מסוג Transitions
- Mismatches מסוג Transversion
- Gaps in sequence 1 – מספר gaps ברצף הראשון בעימוד
- Gaps in sequence 2 – מספר gaps ברצף השני בעימוד

כך למשל עבור העימוד:

```
CCGT----GGATCT
..||| |...|
ATGTAAAAG-GGGT
```

יתקבל המילון:

```
{ 'matches': 4, 'transitions': 2,
  'transversions': 3, 'gaps_1': 4, 'gaps_2': 1 }
```

דווחו בקובץ ex2.docx:

- לכל זוג רצפים, ציינו מי הוא seq1 ומי הוא seq2 בעימוד.
- מה המילון של הספירות שהתקבל.

3. חזרו על התהליך של סעיף 2, הפעם עם פונקציית העימוד globalms עם פרמטרים שתבחרו. כמו בסעיף 2, דווחו בקובץ ex2.docx את התוצאות והשוו לסעיף 2. האם התוצאות שהתקבלו הגיוניות?

הנחיות הגשה:

1. כתבו את התשובות המילוליות בקובץ ex2.docx.
2. ארזו את הקבצים ex2.docx, alignment.py ו-mutations.py ב-ex2.zip.
3. הגישו את הקובץ ex2.zip במערכת המטלות של המכללה.