

תרגיל בית מספר 1

חומר הלימוד לתרגיל: מבוא לפית'ון, מבוא לביולוגיה מולקולרית (רצפי דנ"א, רנ"א, הקוד הגנטי)

שאלה 1:

כתבו פונקציה בשם `is_valid_dna(seq)`. הפונקציה מקבלת מחרוזת ומחזירה `True` אם כל האותיות במחרוזת הן נוקלאוטידים חוקיים (כלומר A או C או G או T). אחרת, הפונקציה מחזירה `False`.
הערה: גם אותיות קטנות (`a\c\g\t`) הן נוקלאוטידים חוקיים.

שאלה 2:

כתבו פונקציה בשם `get_gc_content(seq)` אשר מקבלת רצף דנ"א ומחזירה `GC%`, כלומר אחוז הנוקלאוטידים שהם G או C.
א. הפונקציה מחזירה מספר ממשי.
ב. אם מתקבל רצף שאינו רצף דנ"א חוקי, הפונקציה תחזיר `-1`.

שאלה 3:

כתבו פונקציה בפית'ון בשם `reverse_complement(seq)`. הפונקציה מקבלת מחרוזת (רצף דנ"א) ומחזירה את הרצף שמהווה `reverse complement` עבורה. כך למשל, עבור הפרמטר `'ACGTTG'` הפונקציה תחזיר `'CAACGT'`.
הערות:

א. הפונקציה יכולה לקבל מחרוזת שמורכבת גם מאותיות קטנות וגם מאותיות גדולות, אך היא תחזיר רק אותיות גדולות.
ב. אם מתקבל רצף שאינו רצף דנ"א חוקי, הפונקציה תחזיר מחרוזת ריקה.

שאלה 4: תהליך השעתוק

כתבו פונקציה בשם `get_transcription(seq, strand)`. הפונקציה מקבלת שני פרמטרים:
א. מחרוזת (רצף דנ"א) על הסטרנד החיובי שנרצה לשעתק.
ב. `strand` (מטיפוס `int`: הסטרנד החיובי מיוצג ע"י 1 והשלילי ע"י -1).
הפונקציה מחזירה את הרצף הרנ"א אשר משועתק מהרצף `seq`. אם הסטרנד הוא -1- משמעות הדבר היא שהרצף המקודד הוא על הגדיל המשלים.
הערה: ניתן להניח שהפרמטר של הסטרנד יהיה חוקי (כלומר 1 או -1).

שאלה 5: הקוד הגנטי

הקוד הגנטי זו טבלה שממפה מקודונים לחומצות אמינו. ישנם 64 קודונים ו-20 חומצות אמינו שונות. נרצה לספור עבור כל חומצת אמינו, כמה קודונים שונים מקודדים עבורה. כתבו פונקציה בשם `stats_amino_acids` אשר מחזירה מילון עם ספירות אלה. המפתח במילון יהיה שם של חומצת אמינו (אות אחת) והערך יהיה הספירה המתאימה. ספרו גם כמה קודוני סיום יש, נסמן זאת על ידי המפתח `END`.

הערות:

- עבדו עם הקוד הגנטי שמופיע ב**קישור הזה**.
- הפונקציה לא מקבלת פרמטר, במקום זאת הגדירו את הקוד הגנטי כמשתנה גלובלי בקובץ הפתרון.

שאלה 6: תהליך התרגום

כתבו פונקציה בשם `translate_seq(seq, reading_frame=None)` המתרגמת רצף דנ"א לחלבון. הפונקציה מקבלת מחרוזת שהיא רצף דנ"א (בדקו שהרצף הוא רצף דנ"א חוקי, אם לא, החזירו רשימה ריקה) ומתרגמת את הרצף לחלבון. אם `reading_frame` הוא `None`, הפונקציה מחזירה רשימה של שלושת רצפי החלבון האפשריים שהרצף עשוי לקודד. אחרת, אם `reading_frame` הוא מספר בין 1 ל-3, הפונקציה תחשב את החלבון המתאים כאשר הקריאה מתחילה בעמדה `reading_frame`. במקרה זה הפונקציה תחזיר רשימה באורך אחד, המכילה את רצף החלבון.

הערות:

- תהליך התרגום מתבצע ע"י קריאה של רצפים באורך 3 מהדנ"א ושימוש במילון שהוא הקוד הגנטי שמופיע ב**קישור הזה**.
- הניחו שהרצף שמתקבל הוא על ה `plus strand`.
- שימו לב, אינדקסים בפיית'ון מתחילים מ-0. המשמעות של `reading_frame` שהוא 1 הוא שנתחיל לקרוא את הרצף מתחילתו, מהעמדה הראשונה שלו (0), המשמעות של `reading_frame` שהוא 2 הוא שנתחיל לקרוא את הרצף מהאות השנייה שלו (כלומר אינדקס 1) וכך בהתאם עבור `reading_frame=3`.

שאלה 7: codon bias

כתבו פונקציה בשם `count_codons(seq)` אשר הפונקציה מקבלת מחרוזת שהיא רצף דנ"א (הניחו שהרצף חוקי). הפונקציה סופרת עבור כל קודון שברצף, כמה פעמים הוא מופיע ברצף, ומחזירה מילון עם הספירות המתאימות (המפתח הוא קודון, הערך הוא מספר הופעות). כך למשל עבור הרצף 'ACGTGTTGT' יוחזר המילון:

{ 'ACG': 1, 'TGT': 2 }

(כלומר, יש לספור קודונים ללא חפיפה).

הנחיות הגשה:

1. את התרגיל תוכלו לממש בPyCharm (או בתוכנה אחרת) בקובץ בשם `ex1.py` או ב-google colab בקובץ בשם `ex1.ipynb`. ארזו את הקובץ ב-`ex1.zip`. ציינו בראש הקובץ שם ות"ז.
2. הגישו את הקובץ **ex1.zip** במערכת המטלות של המכללה.