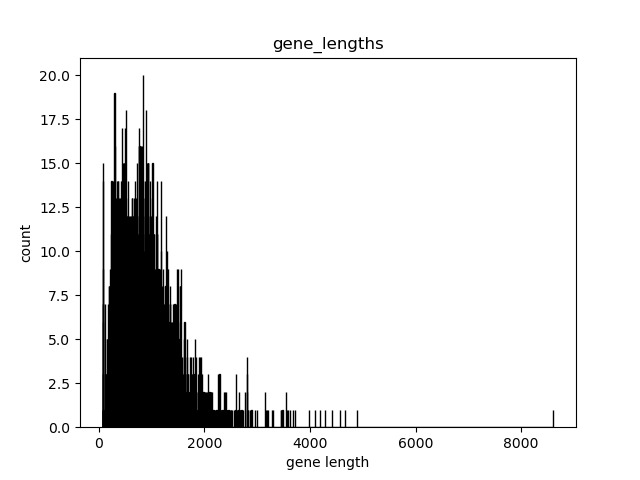
עמרי נהור 211359054

ניצן רפאל מגדל 206763351

יעל בורוכוב 323029637

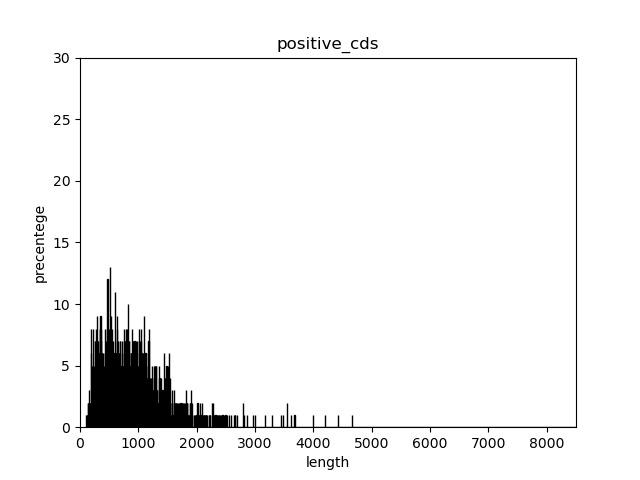
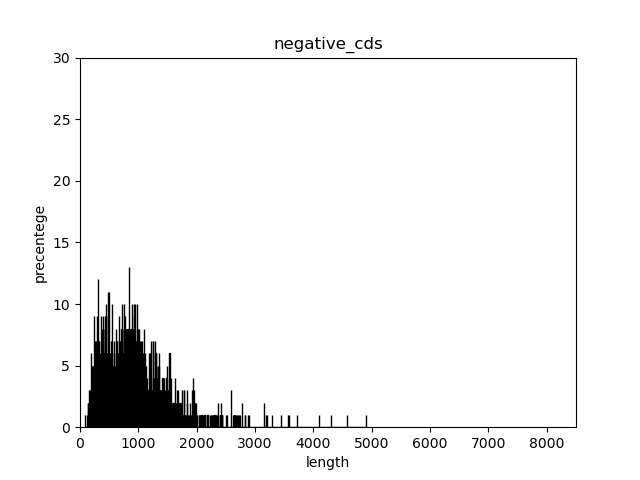
חלק א:

1. {'source': 1, 'gene': 4203, 'CDS': 4108, 'rRNA': 22, 'tRNA': 73, 'misc\_feature': 27}
2. אפיון אורכי הגנים:
3. 
4. positive\_strand: 4072

negative\_strand: 4239

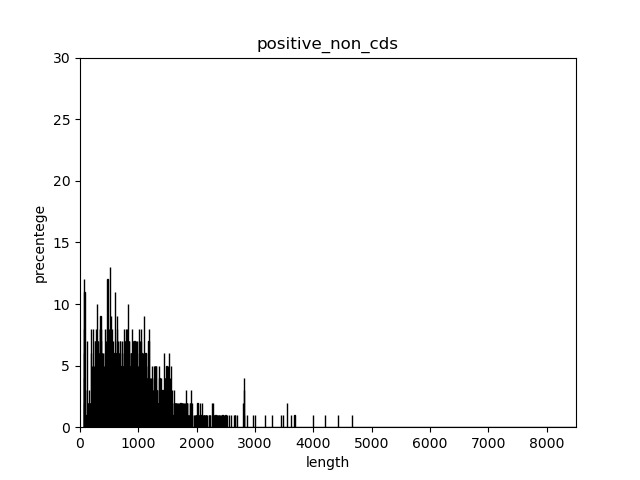
positive\_strand\_with\_cds: 1998

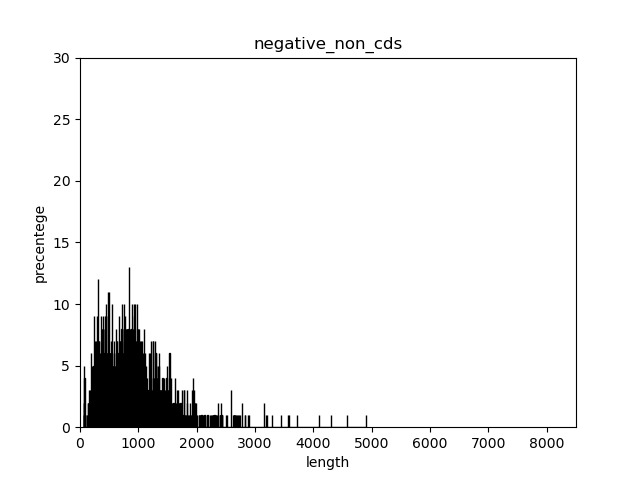
negative\_strand\_cds: 2110



ההתפלגויות דומות. כפי שניתן לראות בסעיף ד' הממוצע וההתפלגות דומים.

1. י

* סעיף ג

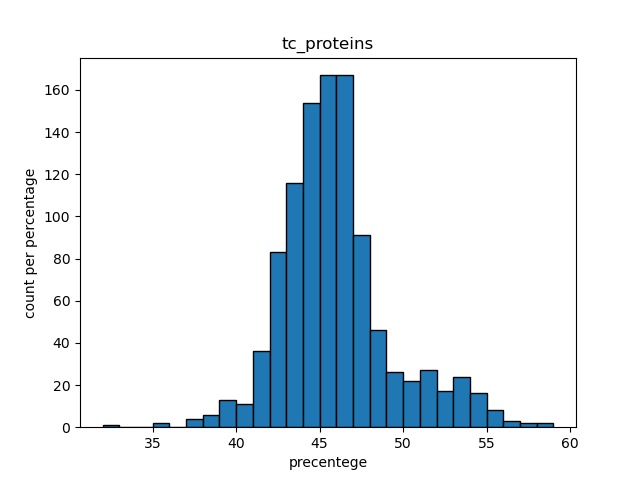


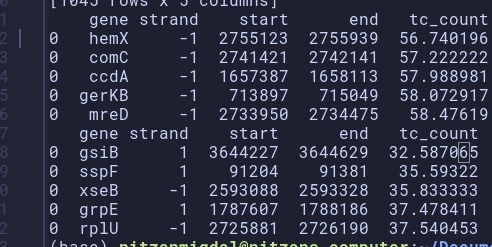
ההתפלגויות דומות. כפי שניתן לראות בסעיף ד' (למטה) הממוצע וההתפלגות דומים.

* סעיף ד:

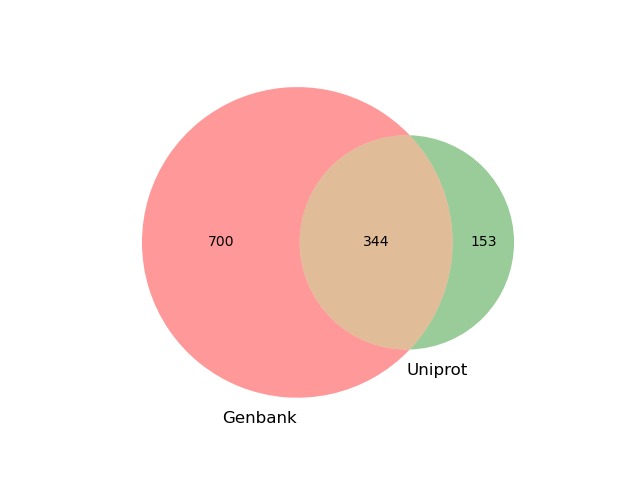
1. ההבדלים בין הסטרנדים: ההתפלגות גדולה יותר בסטרנד החיובי, הממוצע קטן יותר, המינימום והמקסימום גדולים יותר.

ההבדלים בין הגנים שמקודדים לחלבון: ההתפלגות גדולה בשני הסטדנדז באיברים הלא מקושרים לחלבונים הממוצע קטן יותר המקסימום דומה אך המינימום קטן יותר.

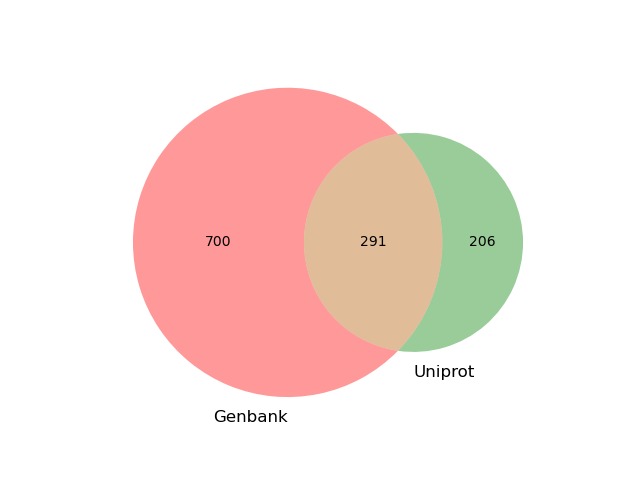
1. חישוב אחוזי CT:
2. tc\_count\_in\_record 50.07290050608635
3. tc\_protein 45.953536552442046
4. בעקבות כך שהגנים המתורגמים לחלבונים הם חלק מהחישוב שחושב בחלק א' ציפינו שהתוצאות יהיו יחסית קרובות לתוצאה הכללית שקיבלנו, התוצאה שקיבלנו הגיונית בעקבות היחס בכמות הגנים המתורגמים לחלבון לעומת כמות הגנים הכללית, שהוא בערך 5 אחוז גדול יותר ובעקבות כך יכלו התוצאות להיות קטנות יותר בעד 5 אחוז.
5. 



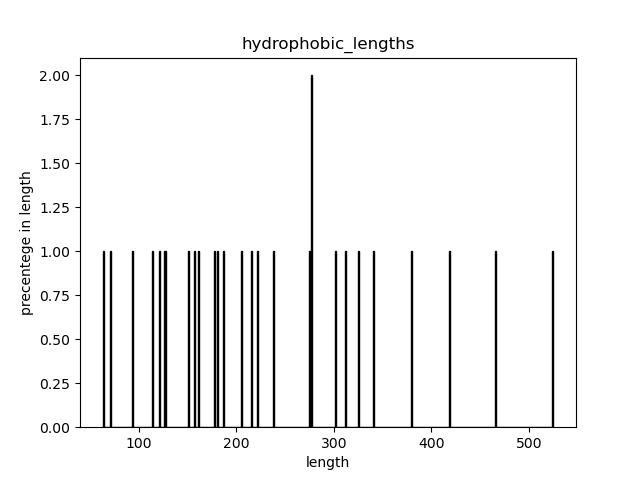
חלק ב'

1. 

זוהי דיאגרמת venn המתארת את ההפרשים בין שני הקבצים על ידי השוואה לפי הtranslation.



זוהי דיאגרמת venn המתארת את ההפרשים בין שני הקבצים על ידי השוואה לפי gene..



{'min': 63, 'max': 525, 'average': 232.10714285714286, 'standard deviation': 116.81986722108111}

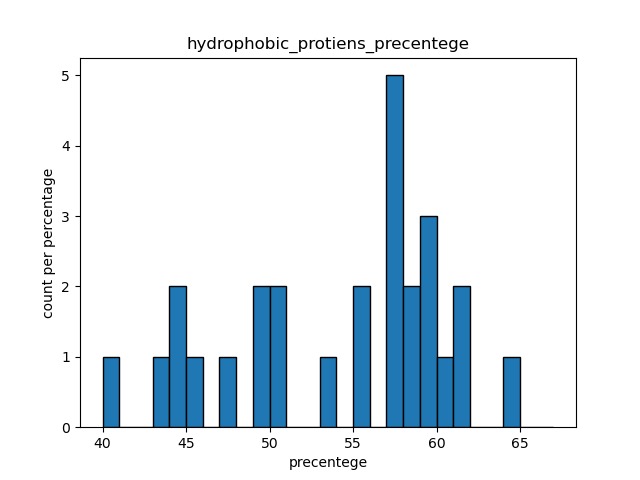
* במחקר : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242586/

נכתב כך :

" Overall, somewhat over 50% of the total amino acid residues in a genome were in sequences within the range of hydrophobicities that were scored."

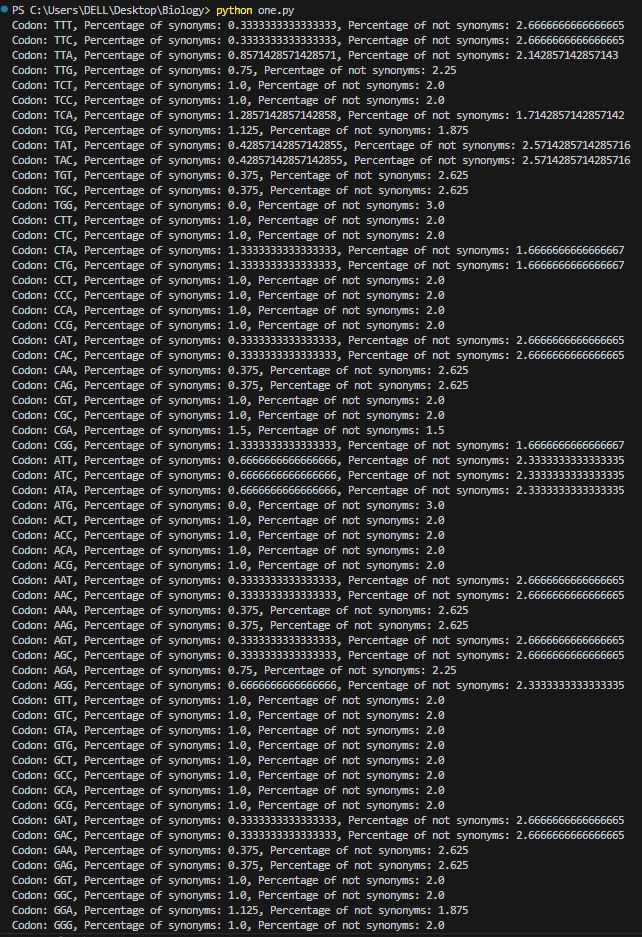
בעקבות כך ציפינו לתוצאות דומות ואכן קיבלנו אותם.

{'min': 40.0, 'max': 67.6056338028169, 'average': 54.536870521160374, 'standard deviation': 6.8414349473320115}

ניתן לראות לפי הנתונים הללו שאחוז חומצות האמינו ההדרופוביות היה 55% בחלקים הטרנסממבריינים.

חלק ג':

1. פתרון:



1. אורך כל גנום:
   1. אפריל 2021: אורך הגנום הוא 29740

פברואר 2024: אורך הגנום הוא 29661

* 1. הגנים שיש בכל אחד מהם:

אפריל 2021:

מספר גנים- 11

גנים שמקודדים לחלבון- 12

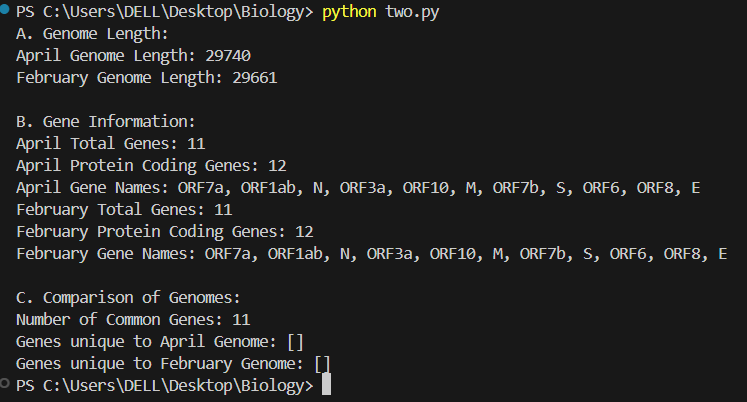
פברואר 2024:

מספר גנים- 11

גנים שמקודדים לחלבון- 12

* 1. לפי הפלט אנו רואים שכל הגנים משותפים. אין גנים שונים.

הגנים: ORF6, S, ORF1ab, ORF3a, ORF10, ORF7b, E, ORF8, N, M, ORF7a



* 1. בחרנו את הגנים הבאים:
     + ORF10-

הגן הזה לא השתנה בכלל לכן הdnds יצא נטרלי.

* + - ORF7a-

הגן הזה לא השתנה בכלל לכן הdnds יצא נטרלי.

* + - ORF3a-

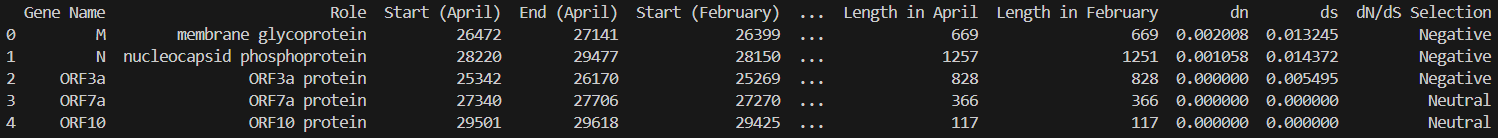
יצא שלילי.

* + - M-
    - N-

לקחנו את הgene חילקנו אותו לקודונים. כל קודון המרנו לחלבון ואז עשינו alignment כדי לבצע התאמה מירבית. הימרנו את הרצף שוב לקודונים וכל "-" הומר ל"---".

שלחנו לפונקציה המובנית של פיתון dnds: ( import CodonSeq, cal\_dn\_ds ) את הרצף הסופי והשמטנו את קודוני העצירה מכיוון שהdnds לא יודע להתמודד עם קודוני העצירה.

למדנו גם בשיעור הראשון שכאשר מגיעים לקודוני העצירה עוצרים.

הטבלה הסופית כולל הדפסה של הdn ds: