נכתב על ידי עידו בן צבי (208627232) שגיא בן שושן (209351147) ועדיאל רוזנפלד (208349019)

**חלק א'**

1.

Number of gene: 4203

Number of CDS: 4108

Number of rRNA: 22

Number of tRNA: 73

Number of misc\_feature: 27

2.  
א. תמונה שמכילה צילום מסך, טקסט, עלילה, קו

התיאור נוצר באופן אוטומטי

ב.  
 Number of genes on plus strand: 2074

Number of genes on minus strand: 2129

Number of protein coding genes on plus strand: 1998

תמונה שמכילה טקסט, צילום מסך, מספר, עלילה

התיאור נוצר באופן אוטומטיNumber of protein coding genes on minus strand: 2110

ג. תמונה שמכילה צילום מסך, תרשים, טקסט, עלילה

התיאור נוצר באופן אוטומטי התפלגויות דומות הממוצע באזור ה 900 בסיסים לגן.

ד.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Std | Mean | Max | Min | Strand | CDS type |
| 565.016 | 900.657 | 8613 | 102 | Plus(1998) |
| 535.217 | 908.709 | 4896 | 81 | Minus(2110) |

ה.

תמונה שמכילה טקסט, צילום מסך, קו, תרשים

התיאור נוצר באופן אוטומטי

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Std | Mean | Max | Min | Strand | tRNA type |
| 4.909 | 77.5 | 94 | 72 | Plus(54) |
| 5.983 | 16.684 | 93 | 71 | Minus(19) |

תמונה שמכילה טקסט, צילום מסך, קו, מספר

התיאור נוצר באופן אוטומטי

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Std | Mean | Max | Min | Strand | rRNA type |
| 1110.503 | 1427.682 | 28004 | 112 | Plus(22) |
| 0 | 0 | 0 | 0 | Minus(0) |

ו.

* ההבדלים בין הסטרנדים:
  + ניתן לראות לגבי גנים שמקודדים לחלבון כי אין הבדל בין הסטרנדים, הסטטיסטיקות מצביעות על כך שאורכי הגנים שמקודדים לחלבון דיי זהים בין סטרנד חיובי לשלילי כמו גם הממוצע וסטיית התקן.
  + לגבי tRNA ניתן לראות כי גם כאן אין הבלד משמעותי בין הסטרנדים, הסטטיסטיקות מצביעות על כך שאורכי הגנים שמקודדים לtRNA דיי זהים בין הסטרנד החיובי לשלילי כמו גם הממוצע וסטיית התקן.
  + לגבי rRNA רואים בבירור שהגנים נמצאים אך ורק על הסטרנד החיובי קיימים 22 גנים כאלו והאורכים שלהם דיי מפוזרים על גבי הגרף מה שמצביע על שונות גדולה (סטיית התקן גדולה מאוד).
* ההבדלי בין גנים שמקודדים לחלבון לבין כאלו שלא מקודדים לחלבון:

ישנם המון גנים שמקודדים לחלבון, למעשה הרוב המוחלט מהגנים של החיידק מקודדים לחלבון, חלק קטן מהגנים מקודדים לRNA ריבוזומאלי ולRNA שליח שהם משתתפים בתהליך ייצור החלבונים בתא.

3. א. מספר נוקלאוטידים מסוג C 965102, מספר נוקלאוטידים מסוג T 1189975, אחוז בסיסים אלו מכול הגנום: 81%.

ב. ממוצע כמות ה CT בגנים המקודדים לחלבון הינו 50% מבסיסי הגן.

ג. בסעיף הקודם מצאנו כי הגנים המקודדים לחלבון נמצאים על שני הגדילים באופן זהה (כמעט חצי חצי) כלומר ישנה כמות דומה של גנים שמקודדים לחלבון על כול סטרנד, כעט מצאנו כי כמות ה C והT של כול הגנים שמקודדים לחלבון היא 50% כלומר חצי מהבסיסים של כול גן הם בסיסים מסוג C T, מכאן שהחצי השני של הבסיסים שלהם מסוגים תואמים G A. ידוע לנו שDNA הוא דו גדילי (ישנם 2 סטרנדים) וגם ידוע לנו שנוקלאוטיד מסוג T בגדיל מסויים תמיד עומד מנגד לנוקלאוטיד מסוג A כמו שG עומד מול C בסטרנד הנגדי.  
לכן אם חצי מהגנים יושבים על סטרנד אחד וחצי מהגנים יושבים על הסטרנד האחר היינו מצפים ליחס בסיסים שכזה.   
המחשה: תמונה שמכילה קו, צילום מסך, צבעוני, טקסט

התיאור נוצר באופן אוטומטי

ד. תמונה שמכילה צילום מסך, עלילה, תרשים

התיאור נוצר באופן אוטומטי

ה.

5 richest CDS genes with the highest C and T nucleotide content:

'name': 'Gene name not available, locus tag: ABC3680', 'start': 3836971, 'end': 3837355, 'strand': -1, 'C\_T\_content\_by\_percent': 0.67  
'name': 'Gene name not available, locus tag: ABC3803', 'start': 3963886, 'end': 3964210, 'strand': -1, 'C\_T\_content\_by\_percent': 0.67  
'name': 'Gene name not available, locus tag: ABC2100', 'start': 2221887, 'end': 2222022, 'strand': -1, 'C\_T\_content\_by\_percent': 0.7  
'name': 'Gene name not available, locus tag: ABC3192', 'start': 3313898, 'end': 3314048, 'strand': -1, 'C\_T\_content\_by\_percent': 0.72  
'name': 'Gene name not available, locus tag: ABC3162', 'start': 3285891, 'end': 3286071, 'strand': -1, 'C\_T\_content\_by\_percent': 0.77

5 poorest CDS genes with the lowest C and T nucleotide content

'name': 'Gene name not available, locus tag: ABC1762', 'start': 1882854, 'end': 1883031, 'strand': 1, 'C\_T\_content\_by\_percent': 0.29  
'name': 'Gene name not available, locus tag: ABC1314', 'start': 1431863, 'end': 1432016, 'strand': 1, 'C\_T\_content\_by\_percent': 0.32  
'name': 'Gene name not available, locus tag: ABC0946', 'start': 1027483, 'end': 1027762, 'strand': 1, 'C\_T\_content\_by\_percent': 0.33  
'name': 'gsiB', 'start': 3644227, 'end': 3644629, 'strand': 1, 'C\_T\_content\_by\_percent': 0.33  
'name': 'Gene name not available, locus tag: ABC0090', 'start': 108960, 'end': 109257, 'strand': 1, 'C\_T\_content\_by\_percent': 0.34