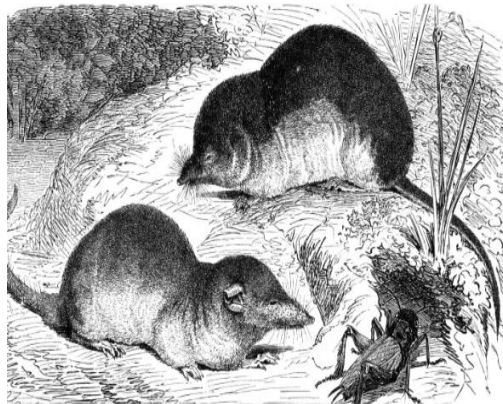




دانشگاه اصفهان
دانشکده مهندسی کامپیوتر

گزارش پروژه

DNA



پدیدآورنده:

محمد امین کیانی

4003613052

دانشجوی کارشناسی، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان،
Aminkianiworkeng@gmail.com

استاد: جناب آقای دکتر نقش نیلچی

نیمسال اول تحصیلی 1403-04

فهرست مطالب

3	مستندات
3	1- مسئله و تحلیل کلی آن:
6	2- توضیح کد برنامه:
10	3- تحلیل نتایج:
14	4- خروجی‌ها و نمودارها:
19	5- مراجع:

مستندات

1- مسئله و تحلیل کلی آن:

DNA (دئوکسی ریبونوکلئیک اسید) مولکولی است که اطلاعات ژنتیکی موجودات زنده را ذخیره می‌کند. این مولکول از واحدهایی به نام **نوکلئوتید** تشکیل شده است. هر نوکلئوتید شامل سه بخش اصلی است:

1. یک گروه فسفات.
2. یک قند پنج‌کربنی (دئوکسی‌ریبوز).
3. یک باز نیتروژنی که می‌تواند یکی از چهار نوع زیر باشد :

- آدنین (A)
- تیمین (T)
- گوانین (G)
- سیتوزین (C)

دی‌ان‌ای از دو رشته مکمل تشکیل شده است که بازهای نیتروژنی آن‌ها با پیوندهای هیدروژنی به هم متصل هستند. بازهای A و T با دو پیوند هیدروژنی و بازهای G و C با سه پیوند هیدروژنی جفت می‌شوند.

ژن، واحد اصلی اطلاعات ژنتیکی در DNA است. هر ژن دنباله‌ای از نوکلئوتیدهاست که اطلاعات لازم برای ساخت پروتئین‌ها را فراهم می‌کند. ژن‌ها به صورت کدهای سه‌تایی نوکلئوتیدها (کدون‌ها) خوانده می‌شوند و هر کدون یک اسید آمینه خاص را مشخص می‌کند.

پایداری و مقاومت حرارتی DNA

پایداری و مقاومت حرارتی DNA به ساختار شیمیایی آن وابسته است. یکی از مهمترین معیارهای بررسی این ویژگی‌ها، **درصد GC** است:

- **GC Content** به درصد بازهای گوانین (G) و سیتوزین (C) در دنباله DNA اشاره دارد.

- بازهای G و C به دلیل سه پیوند هیدروژنی که دارند، پایداری بیشتری نسبت به بازهای A و T (که فقط دو پیوند هیدروژنی دارند) ایجاد می‌کنند.
- **GC بالا** (بیش از 50%) نشان‌دهنده پایداری و مقاومت حرارتی بالاتر است.

معرفی موجود *Sorex roboratus* :

گونه‌ای از پستانداران کوچک و متعلق به خانواده **حشره‌خواران (Soricidae)** است. این موجودات به دلیل اندازه کوچک، زیستگاه‌های خاص، و نقش کلیدی در اکوسیستم‌های زمینی مورد توجه زیست‌شناسان هستند.

ژنوم میتوکندری:

ژنوم میتوکندری موجودات شامل اطلاعات ژنتیکی است که برای عملکرد میتوکندری‌ها لازم است. میتوکندری، اندامکی است که مسئول تولید انرژی سلول از طریق فرآیند تنفس سلولی می‌باشد. در بیشتر جانداران، ژنوم میتوکندری ساختاری حلقوی داشته و شامل ژن‌های زیر است:

- ژن‌های کدکننده پروتئین‌های مرتبط با زنجیره انتقال الکترون.
- ژن‌های کدکننده tRNA و rRNA برای سنتز پروتئین.

ژنوم میتوکندری به دلیل اندازه کوچک و سرعت تکامل سریع، یکی از بهترین منابع برای مطالعات ژنتیکی و فیلوژنتیکی است.

The image displays two side-by-side windows of a sequence editor. The left window shows a raw DNA sequence with line numbers from 1 to 2401. The right window shows the same sequence with various annotations. The annotations include:

- LOCUS:** NC_034808 17153 bp DNA circular MAP 03-APR-2023
- DEFINITION:** Sorex roboratus mitochondrion, complete genome.
- ACCESSION:** NC_034808
- VERSION:** NC_034808.1
- DBLINK:** BioProject: PRJNA927338
- KEYWORDS:** RefSeq.
- SOURCE:** mitochondrion Sorex roboratus (flat-skulled shrew)
- ORGANISM:** Sorex roboratus
- REFERENCE:** Liu, Z. Sequencing and analysis of the complete mitochondrial genome of flat-skulled shrew (Sorex roboratus) from China
- AUTHORS:** Liu, Z.
- JOURNAL:** Submitted (11-APR-2017) Mudanjiang Normal University, college of Life Science and Technology, wenhua street No. 191, mudanjiang, heilongjiang, China
- CONSTRN:** NCBI Genome Project
- TITLE:** Direct Submission
- JOURNAL:** Submitted (24-MAY-2017) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
- REFERENCE:** Liu, Z.
- AUTHORS:** Liu, Z.
- JOURNAL:** Submitted (11-APR-2017) Mudanjiang Normal University, college of Life Science and Technology, wenhua street No. 191, mudanjiang, heilongjiang, China
- COMMENT:** This record has been curated by NCBI staff. The reference sequence is identical to KY930906.
- FEATURES:** source 1..17153 /organism="Sorex roboratus" /organism_loc="mitochondrion" /mol_type="genomic DNA" /db_xref="taxon:62288" 1..69 /locus_tag="SRRMS_met01"
- gene:**

توضیحات دنباله DNA

ژنوم میتوکندری *Sorex roboratus* شامل حدود 17,000 جفت باز (bp) است که تمام ژن‌های مورد نیاز برای عملکرد میتوکندری را در بر می‌گیرد. دنباله DNA شامل نواحی کدکننده ژن‌ها و نواحی غیرکدکننده (کنترلی) است.

اهمیت درصد GC در ژنوم میتوکندری

- درصد GC به دلیل تاثیر آن بر پایداری ساختار DNA از اهمیت ویژه‌ای برخوردار است:
- بازهای گوانین (G) و سیتوزین (C) دارای سه پیوند هیدروژنی هستند که باعث افزایش پایداری حرارتی DNA می‌شود.
- در ژنوم میتوکندری، درصد GC معمولاً پایین‌تر از ژنوم هسته‌ای است و این به دلیل ساختار و عملکرد متفاوت آن است.

مراحل انجام پروژه:

1- دریافت دنباله DNA از GenBank

GenBank یک پایگاه داده زیستی شامل اطلاعات مربوط به توالی‌های DNA است که برای دریافت دنباله باید مراحل زیر را طی کنیم:

1. به وبسایت GenBank مراجعه می‌کنیم.
2. ژن مورد نظر خود را جستجو می‌کنیم (*Sorex roboratus mitochondrion*).
3. دنباله ژن را انتخاب کرده و فایل آن را با فرمت **gb** دانلود می‌کنیم. ژنوم میتوکندری *Sorex roboratus* را می‌توان از پایگاه داده GenBank دریافت کرد و فایل با فرمت **gb** حاوی اطلاعات کامل ژنوم است.

2- تحلیل دنباله DNA با Python

به کمک زبان پایتون انجام موارد زیر را انجام می‌دهیم:

1. شمارش نوکلئوتیدها

نوکلئوتیدهای ساده شامل A ، T ، G و C هستند. برنامه تعداد هر یک از این نوکلئوتیدها را محاسبه کرده و درصد آنها را نسبت به کل دنباله گزارش می‌کند.

2. شمارش دی‌نوکلئوتیدها و تری‌نوکلئوتیدها

- دی‌نوکلئوتید: دنباله‌های دو حرفی مانند AT ، GC و غیره.
 - تری‌نوکلئوتید: دنباله‌های سه حرفی مانند ATG ، CGT و غیره.
- برنامه تعداد این کلمات را محاسبه کرده و درصد آنها را گزارش می‌کند.

3. محاسبه درصد GC

فرمول GC Content درصد بازهای G و C در کل دنباله است که به صورت زیر محاسبه می‌شود:

$$\text{GC Content (\%)} = \frac{\text{Count of G} + \text{Count of C}}{\text{Total Bases}} \times 100$$

این معیار برای تحلیل پایداری و مقاومت حرارتی استفاده می‌شود که دو حالت دارد:

- GC بالا (بیش از 50%): پایداری و مقاومت حرارتی بیشتر به دلیل وجود 3 پیوند هیدروژنی.
- GC پایین (کمتر از 50%): پایداری کمتر.

2- توضیح کد برنامه:

در این بخش، کدی که با استفاده از زبان Python نوشته‌ایم، تحلیل دنباله DNA را انجام می‌دهد. هر بخش کد به شرح زیر می‌باشد:

```
!pip install Bio

from Bio import SeqIO
from collections import Counter
import pandas as pd
```

- **کتابخانه Bio** : پس از نصب، آن را برای خواندن فایل GenBank که دنباله DNA را شامل می‌شود، استفاده می‌کنیم.
- **کتابخانه Counter** : برای شمارش فراوانی نوکلئوتیدها و کلمات دو و سه حرفی.
- **کتابخانه Pandas** : برای ساخت و نمایش جداول نتایج.

تابع اصلی تحلیل دنباله DNA :

```
def DNA_Analyzer(GB_file_path):
    # load from GenBank
    record = SeqIO.read(GB_file_path, "genbank")
    dna_sequence = str(record.seq).upper()
```

- `file_path` : مسیر فایل GenBank شامل دنباله DNA
- `SeqIO.read` : دنباله DNA را از فایل می‌خواند.
- `dna_sequence` : دنباله DNA به صورت رشته بزرگ (uppercase)

بررسی طول دنباله

```
if len(dna_sequence) < 2000:
    raise ValueError("طول دنباله دی‌ان‌ای باید حداقل 2000 نوکلئوتید باشد.")
```

- طبق صورت سوال پروژه بررسی می‌کند که دنباله حداقل 2000 نوکلئوتید باشد و اگر کمتر باشد، خطا می‌دهد زیرا دنباله‌ی مورد بررسی باید طویل باشد.

شمارش نوکلئوتیدهای ساده

```
# شمارش نوکلئوتیدهای ساده
simple_counts = Counter(dna_sequence)
simple_percentages = {nuc: count / len(dna_sequence) * 100 for nuc,
count in simple_counts.items() }
```

- `Counter` : تعداد هر نوکلئوتید را می‌شمارد.
- سپس درصد هر نوکلئوتید محاسبه می‌شود.

شمارش دی‌نوکلئوتیدها

```
# شمارش دی‌نوکلئوتیدها
di_nucleotides = [dna_sequence[i:i+2] for i in range(len(dna_sequence)
- 1)]
di_counts = Counter(di_nucleotides)
di_percentages = {dinuc: count / len(di_nucleotides) * 100 for dinuc,
count in di_counts.items() }
```

- تمام ترکیبات دو حرفی در دنباله استخراج می‌شوند.
- سپس Counter تعداد هر ترکیب دو حرفی را می‌شمارد.

شمارش تری‌نوکلئوتیدها

```
# شمارش تری‌نوکلئوتیدها
tri_nucleotides = [dna_sequence[i:i+3] for i in
range(len(dna_sequence) - 2)]
tri_counts = Counter(tri_nucleotides)
tri_percentages = {trinuc: count / len(tri_nucleotides) * 100 for
trinuc, count in tri_counts.items() }
```

- تمام ترکیبات سه حرفی استخراج و شمارش می‌شوند.

ساخت جداول

```
# ساخت جداول
simple_table = pd.DataFrame({
    "Nucleotide": list(simple_counts.keys()),
    "Count": list(simple_counts.values()),
    "Percentage": list(simple_percentages.values())
})

di_table = pd.DataFrame({
    "Di-nucleotide": list(di_counts.keys()),
    "Count": list(di_counts.values()),
    "Percentage": list(di_percentages.values())
})

tri_table = pd.DataFrame({
    "Tri-nucleotide": list(tri_counts.keys()),
```



```

    "Count": list(tri_counts.values()),
    "Percentage": list(tri_percentages.values())
})

```

• جدولی شامل تعداد و درصد نوکلئوتیدها را مطابق فرمت خواسته شده، می‌سازد.

محاسبه درصد GC

```

# محاسبه GC
gc_content = (simple_counts.get('G', 0) + simple_counts.get('C', 0)) /
len(dna_sequence) * 100

return simple_table, di_table, tri_table, gc_content

```

- تعداد بازهای G و C جمع می‌شود.
- درصد GC محاسبه می‌شود.
- تابع در نهایت تمامی خروجی‌ها را برمی‌گرداند.

بکارگیری تابع و نمایش خروجی

```

•
• # استفاده از تابع
• GB_file_path = "sequence.gb"
• simple_table, di_table, tri_table, gc_content =
  DNA_Analyzer(GB_file_path)
•
• # نمایش نتایج
• print("جدول نوکلئوتیدهای ساده:")
• print(simple_table)
•
• print("\nجدول دی‌نوکلئوتیدها:")
• print(di_table)
•
• print("\nجدول تری‌نوکلئوتیدها:")
• print(tri_table)
•
• print(f"\nدرصد GC: {gc_content:.2f}%") #High GC content (>50%)
  indicates greater stability and heat resistance.

```

3- تحلیل نتایج:

تحلیل ژنوم *Sorex roboratus*

ژنوم میتوکندری *Sorex roboratus* اطلاعات ارزشمندی برای تحلیل‌های ژنتیکی و فیلوژنتیکی فراهم می‌کند. تحلیل درصد GC و ترکیبات نوکلئوتیدی، ویژگی‌های پایداری و مقاومت حرارتی این ژنوم را برجسته می‌کند. نتایج حاصل از این پروژه می‌تواند در تحقیقات زیست‌شناسی و پزشکی کاربرد داشته باشد.

تذکر: عبارت **bps** مخفف **base pairs** (جفت بازها) است و به تعداد بازهای آلی A، T، G، C در یک توالی DNA اشاره دارد.

1- جدول نوکلئوتیدهای ساده

خروجی جدول نشان می‌دهد که چه میزان از ژنوم توسط هر نوکلئوتید A، T، G، C تشکیل شده است.

Nucleotide	Count	Percentage
G	2245	13.09%
T	5047	29.42%
C	4182	24.38%
A	5679	33.11%

A و T فراوان‌ترین نوکلئوتیدها هستند:

نوکلئوتیدهای آدنین (A) و تیمین (T) با مجموع 62.53% بیشترین بخش از توالی را تشکیل داده‌اند. این موضوع نشان‌دهنده درصد پایین GC در این ژنوم است.

C و G درصد کمتری دارند:

نوکلئوتیدهای گوانین (G) و سیتوزین (C) جمعاً 37.47% از توالی را تشکیل می‌دهند. این مقدار کمتر نسبت به A و T نشان‌دهنده محتوای GC پایین‌تر و پایداری حرارتی متوسط ژنوم است.

2- جدول دی نوکلئوتیدها و تری نوکلئوتیدها

این جداول توزیع ترکیبات دو و سه حرفی را نشان می دهد که اطلاعات مهمی درباره ساختار ژنومی فراهم می کند.

جدول دی نوکلئوتیدها

Di-nucleotide	Count	Percentage
GT	589	3.43%
TT	1479	8.62%
TC	1133	6.61%
CA	1397	8.14%
AT	1691	9.86%
...

فراوانی بالای دی نوکلئوتیدهای AT و TA :

دو نوکلئوتید (9.86%) AT و (10.74%) TA بیشترین فراوانی را دارند. این موضوع با درصد بالای A و T در ژنوم همخوانی دارد و نشان دهنده وجود مناطق غنی از A و T در توالی است.

GC کمتر از AT :

درصد دی نوکلئوتیدهای (3.28%) GC نسبت به AT بسیار پایین تر است. این نشان دهنده توزیع نامتقارن بازها و محتوای پایین GC است.

فراوانی دی نوکلئوتیدهای تکراری AA و TT :

وجود درصد بالای (10.31%) AA و (8.62%) TT نشان می دهد که ژنوم شامل مناطق تکراری است که ممکن است به تنظیم بیان ژن کمک کند.

جدول تری نوکلئوتیدها

Tri-nucleotide	Count	Percentage
GTT	161	0.94%
TTC	369	2.15%
TCA	369	2.15%
CAT	411	2.40%
ATG	214	1.25%
...

فراوانی متعادل تری نوکلئوتیدها:

ترکیبات مختلف به طور متعادل توزیع شده اند. این تنوع می تواند نشان دهنده عملکردهای مختلف ژنوم و توالی های رمزگذاری کننده برای پروتئین های متنوع باشد.

ترکیبات آغازگر کدون (ATG):

ATG که کدون آغازگر بسیاری از پروتئین هاست، با درصد 1.25% به فراوانی در ژنوم وجود دارد و این منطقی است، زیرا این ترکیب نشان دهنده مناطق کدکننده ژن است.

توالی های غنی از AT:

بسیاری از تری نوکلئوتیدهای غالب مانند GTT و CAT شامل A یا T هستند که نشان دهنده غلبه این بازها در ساختار ژنوم است.

3- درصد GC

درصد GC برای ژنوم میتوکندری *Sorex roboratus* تخمین زده می شود. این مقدار معمولاً کمتر از 50% است و نشان دهنده پایداری متوسط در شرایط دمایی مختلف است. درصد GC پایین (کمتر از 50%) نشان می دهد که توالی ژنوم عمدتاً غنی از بازهای A و T است. بازهای GC به دلیل وجود سه پیوند هیدروژنی، پایداری بیشتری

نسبت به بازهای AT (که دو پیوند هیدروژنی دارند) به DNA می‌دهند. بنابراین، ژنومی با محتوای GC پایین‌تر پایداری حرارتی کمتری دارد.

چرا به این مقدار رسیدیم؟

- ژنوم میتوکندری معمولاً درصد GC کمتری نسبت به ژنوم هسته‌ای دارد. دلیل این امر وابستگی به عملکرد خاص میتوکندری و همچنین محدودیت‌های تکاملی مرتبط با آن است.

4- پایداری و مقاومت در برابر گرما

1. پایداری ژن:

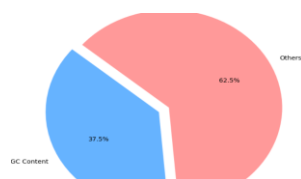
- محتوای GC پایین نشان‌دهنده پایداری متوسط است. چنین ژنومی در برابر تغییرات محیطی و شیمیایی پایداری کمتری نسبت به توالی‌هایی با محتوای GC بالا دارد.

2. مقاومت در برابر گرما:

- DNA با محتوای GC بالا، مقاومت حرارتی بیشتری دارد، زیرا پیوندهای هیدروژنی بیشتری در بازهای G و C وجود دارد.
- برای ژنوم *Sorex roboratus* با $GC = 37.47\%$ ، مقاومت در برابر گرما در سطح متوسط قرار دارد. احتمالاً در دماهای بالای معمول، DNA این ژنوم ناپایدار می‌شود.

5- نتیجه‌گیری کلی

1. ژنوم میتوکندری *Sorex roboratus* غنی از A و T است که نشان‌دهنده مناطق خاص ژنتیکی و تکاملی این موجود است.
2. درصد GC پایین (37.47%) نشان می‌دهد که این ژنوم مقاومت حرارتی بالایی ندارد و احتمالاً در دماهای بالا ناپایدار می‌شود.
3. این اطلاعات می‌تواند در مطالعات تکاملی و همچنین بررسی ویژگی‌های زیستی این موجود مورد استفاده قرار گیرد.



4- خروجی ها و نمودارها:

Mono-Nucleotides:																
	in bps				in %											
A	5679	33.107911														
T	5047	29.423424														
C	4182	24.380575														
G	2245	13.088090														
Di-Nucleotides (in bps):																
	A	T	C	G												
A	1768	1691	1379	841												
T	1842	1479	1133	593												
C	1397	1288	1108	388												
G	672	589	562	422												
Di-Nucleotides (in %)																
	A	T	C	G												
A	10.307836	9.858909	8.039879	4.903218												
T	10.739272	8.622901	6.605644	3.457323												
C	8.144823	7.509328	6.459888	2.262127												
G	3.917910	3.434002	3.276586	2.460354												
Tri-Nucleotide (in bps):																
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	\
A	543	556	431	238	518	575	384	214	486	417	315	161	189	213	266	
T	553	579	422	288	551	383	369	176	369	375	294	94	234	134	118	
C	435	411	366	185	524	360	268	136	333	347	325	103	117	134	79	
G	237	145	160	130	249	161	112	67	209	149	174	30	132	107	99	
GG																
A	173															
T	107															

	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	\
A	543	556	431	238	518	575	384	214	486	417	315	161	189	213	266	
T	553	579	422	288	551	383	369	176	369	375	294	94	234	134	118	
C	435	411	366	185	524	360	268	136	333	347	325	103	117	134	79	
G	237	145	160	130	249	161	112	67	209	149	174	30	132	107	99	
GG																
A	173															
T	107															
C	58															
G	84															
Tri-Nucleotide (in %):																
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	\
A	3.165996	3.241793	2.512973	1.387674	3.020232	3.352574	2.238937									
T	3.224302	3.375896	2.460498	1.679202	3.212641	2.233106	2.151478									
C	2.536295	2.396362	2.133986	1.078654	3.055215	2.099003	1.562591									
G	1.381844	0.845432	0.932890	0.757973	1.451810	0.938721	0.653023									
	TG	CA	CT	CC	CG	GA	GT	\								
A	1.247741	2.833654	2.431345	1.836628	0.938721	1.101977	1.241910									
T	1.026179	2.151478	2.186461	1.714186	0.548073	1.364352	0.781296									
C	0.792957	1.941578	2.023206	1.894933	0.600548	0.682176	0.781296									
G	0.390648	1.218588	0.868754	1.014518	0.174917	0.769634	0.623870									
	GC	GG														
A	1.550930	1.008688														
T	0.688007	0.623870														
C	0.460615	0.338173														
G	0.577226	0.489767														
Percentage of GC: 37.47%																

(درصد و bps) جدول نوکلئوتیدی ساده		
	bps	Percentage
A	5679	33.1079
T	5047	29.4234
C	4182	24.3806
G	2245	13.0881

جدول دی‌نوکلئوتیدها (bps):				
	A	T	C	G
A	1768	1691	1379	841
T	1842	1479	1133	593
C	1397	1288	1108	388
G	672	589	562	422

جدول دی‌نوکلئوتیدها (درصد):				
	A	T	C	G
A	10.3078	9.85891	8.03988	4.90322
T	10.7393	8.6229	6.60564	3.45732
C	8.14482	7.50933	6.45989	2.26213
G	3.91791	3.434	3.27659	2.46035

جدول تری‌نوکلئوتیدها (bps):																
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
A	543	556	431	238	518	575	384	214	486	417	315	161	189	213	266	173
T	553	579	422	288	551	383	369	176	369	375	294	94	234	134	118	107
C	435	411	366	185	524	360	268	136	333	347	325	103	117	134	79	58
G	237	145	160	130	249	161	112	67	209	149	174	30	132	107	99	84

جدول تری‌نوکلئوتیدها (درصد):																
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
A	10.3078	9.85891	8.03988	4.90322												
T	10.7393	8.6229	6.60564	3.45732												
C	8.14482	7.50933	6.45989	2.26213												
G	3.91791	3.434	3.27659	2.46035												

جدول تری‌نوکلئوتیدها (bps):																
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
A	543	556	431	238	518	575	384	214	486	417	315	161	189	213	266	173
T	553	579	422	288	551	383	369	176	369	375	294	94	234	134	118	107
C	435	411	366	185	524	360	268	136	333	347	325	103	117	134	79	58
G	237	145	160	130	249	161	112	67	209	149	174	30	132	107	99	84

جدول تری‌نوکلئوتیدها (درصد):																
	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
A	3.166	3.24179	2.51297	1.38767	3.02023	3.35257	2.23894	1.24774	2.83365	2.43135	1.83663	0.938721	1.10198	1.24191		
T	3.2243	3.3759	2.4605	1.6792	3.21264	2.23311	2.15148	1.02618	2.15148	2.18646	1.71419	0.548073	1.36435	0.781296		
C	2.5363	2.39636	2.13399	1.07865	3.05522	2.099	1.56259	0.792957	1.94158	2.02321	1.89493	0.600548	0.682176	0.781296		
G	1.38184	0.845432	0.93289	0.757973	1.45181	0.938721	0.653023	0.390648	1.21859	0.868754	1.01452	0.174917	0.769634	0.62387		

درصد GC: 37.47%

Mono-Nucleotides (bps and %)

	bps	Percentage
A	5679.0	33.107911152568064
T	5047.0	29.42342447385297
C	4182.0	24.380574826560952
G	2245.0	13.088089547018015

Di-Nucleotides (bps)

	A	T	C	G
A	1768	1691	1379	841
T	1842	1479	1133	593
C	1397	1288	1108	388
G	672	589	562	422

Di-Nucleotides (%)

	A	T	C	G
A	10.307835820895523	9.858908582089553	8.039878731343283	4.903218283582089
T	10.739272388059701	8.622901119402986	6.605643656716417	3.4573227611940296
C	8.14482276119403	7.509328358208956	6.459888059701492	2.262126865671642
G	3.9179104477611943	3.434001865671642	3.276585820895522	2.4603544776119404

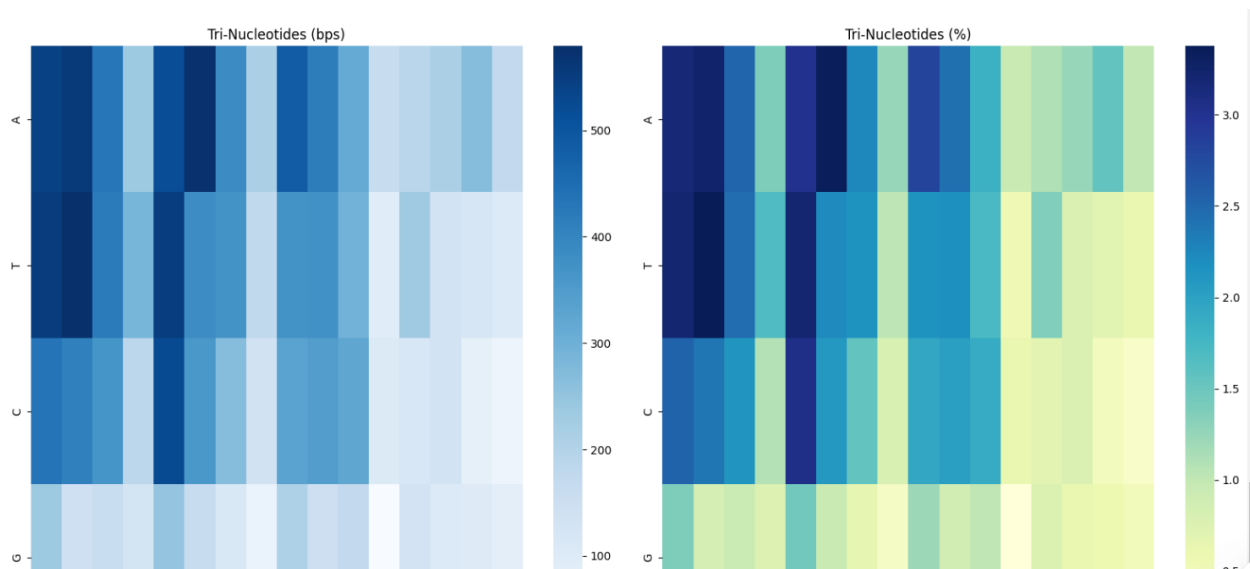
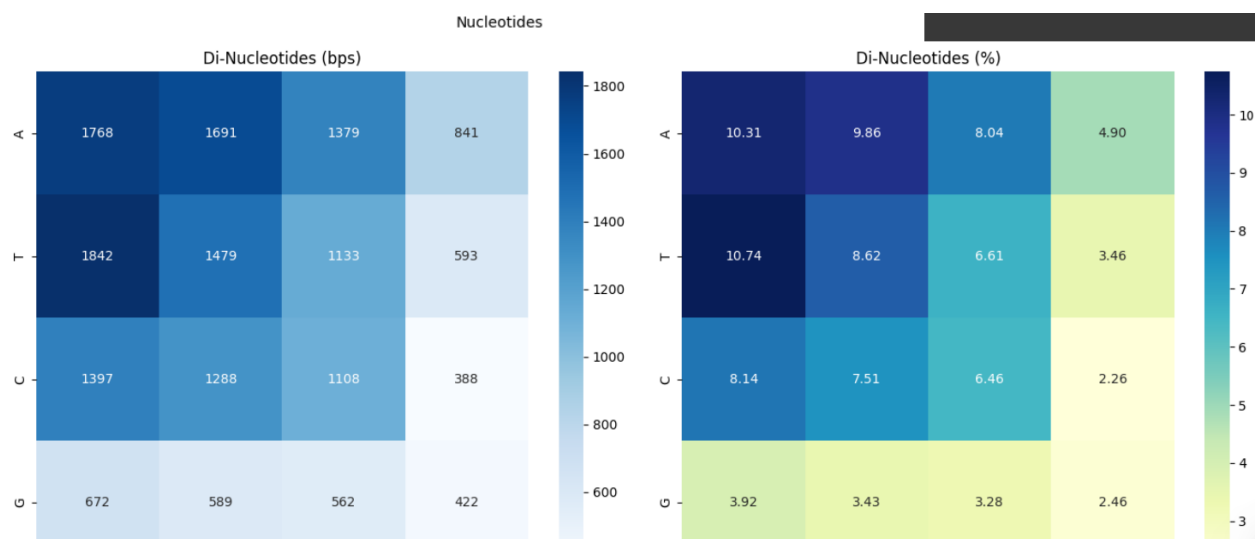
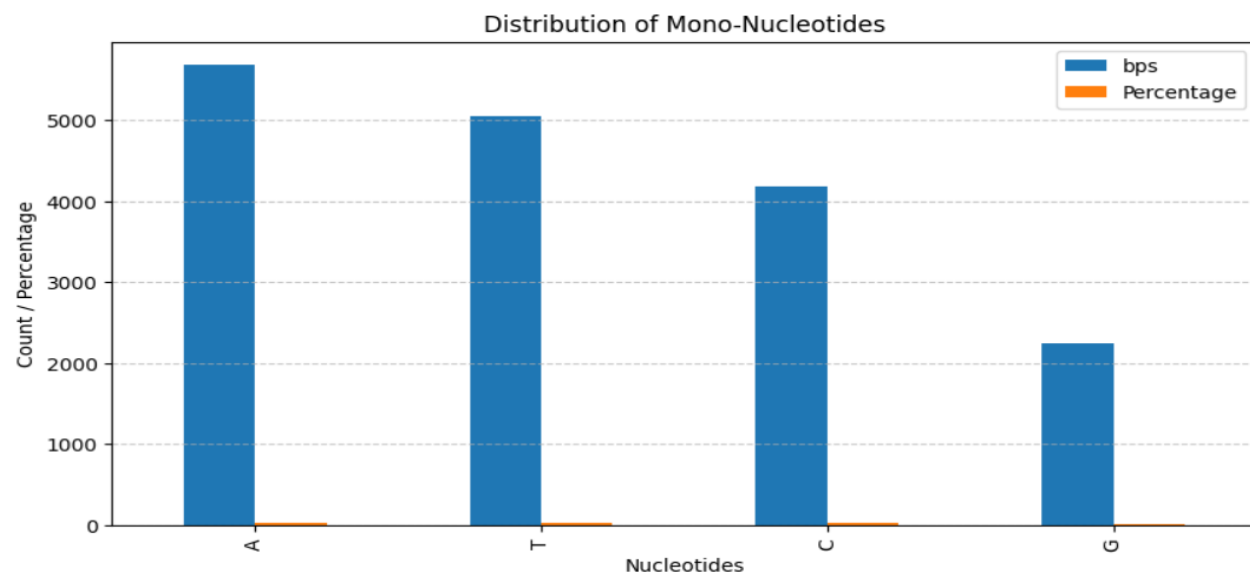
Tri-Nucleotides (bps)

	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
A	543	556	431	238	518	575	384	214	486	417	315	161	189	213	266	173
T	553	579	422	288	551	383	369	176	369	375	294	94	234	134	118	107
C	435	411	366	185	524	360	268	136	333	347	325	103	117	134	79	58
G	237	145	160	130	249	161	112	67	209	149	174	30	132	107	99	84

Tri-Nucleotides (%)

	TT	TC	TG	CA	CT	CC	
3.3525741939245526	2.2389365051600487	1.2477406565214855	2.833654014343187	2.4313451110722406	1.8366276018891028	0.9387207742988747	0.9387207742988747
2.2331059413445282	2.1514780479272346	1.026179231531689	2.1514780479272346	2.1864614308203603	1.7141857617631626	0.5411111111111111	0.5411111111111111
2.0990029735875457	1.5625911025596173	0.7929566789108506	1.94157775056848	2.0232056439857735	1.8949332400443122	0.6000000000000001	0.6000000000000001
0.9387207742988747	0.6530231473383477	0.3906477756399043	1.2185878374438808	0.8687540085126231	1.0145181039006472	0.1746031746031746	0.1746031746031746

```
11] 1 import matplotlib.pyplot as plt
```



- [1] <https://www.ncbi.nlm.nih.gov/>
- [2] https://www.ncbi.nlm.nih.gov/nuccore/NC_034808.1
کد را با دنباله‌ی بالا به عنوان انتخاب اصلی اجرا کردیم و به 37% رسیدیم که یعنی ناپایدارتر است.
- [3] <https://www.ncbi.nlm.nih.gov/nuccore/AL035632.2>
کد را با دنباله‌ی بالا به عنوان تست دیگر مثال‌ها نیز اجرا کردیم و به 53% رسیدیم که یعنی این دنباله پایدارتر است.
- [4] <https://colab.research.google.com/>
- [5] <https://www.researchgate.net/publication/318192460 Sequencing and analysis of the complete mitochondrial genome of flat-skulled shrew Sorex roboratus from China>
- [6] <https://github.com/M-Amin-Kiani/BioInfo/tree/main/research>
لینک گیت‌هاب بالا حاوی کدها و ژن بانک مربوطه برای اجرای کد پروژه به همراه تمامی گزارش‌های این درس است.

Sorex roboratus mitochondrion, complete genome

NCBI Reference Sequence: NC_034808.1

GenBank FASTA

