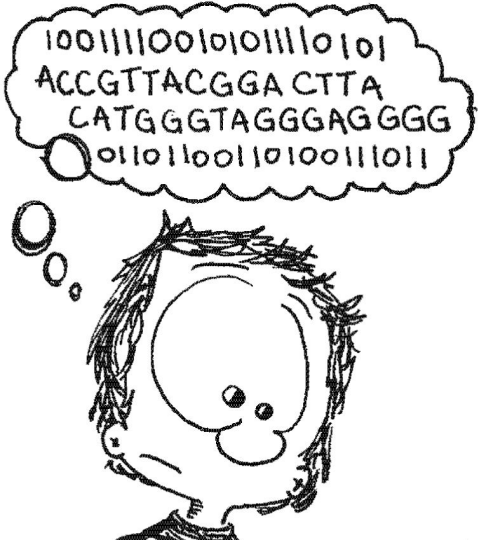


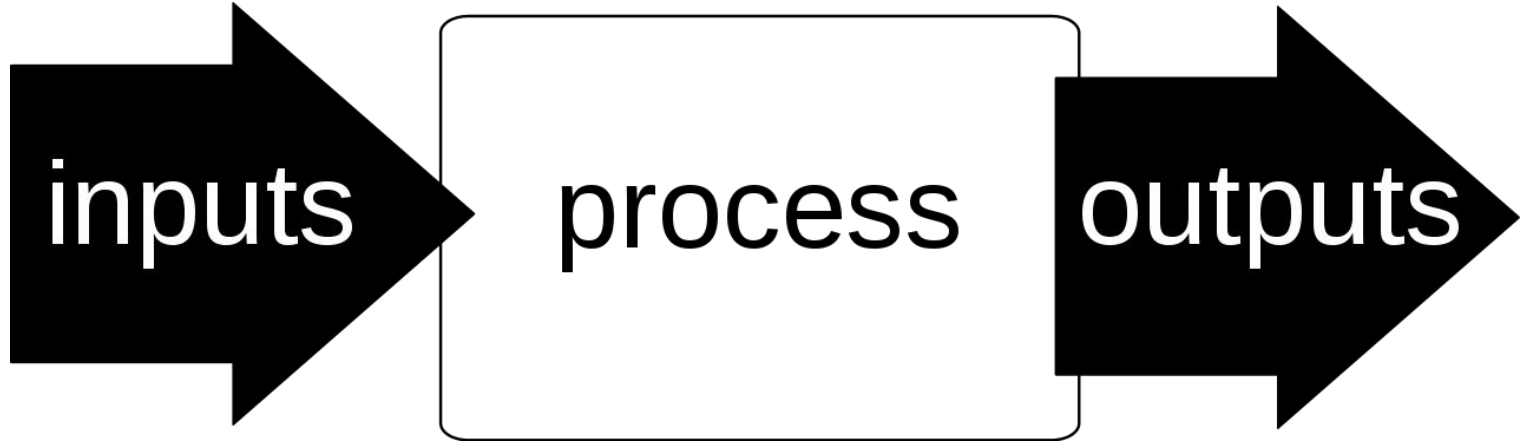
Hesaplamalı Biyoloji ve Biyoinformatiğe Giriş: Uygulamalar



Teorik: Barış Özüdoğru
Pratik: Işın Altınkaya

Hacettepe Üniversitesi, Biyoloji Bölümü





Girdiler -> Süreç -> Çıktılar

Problem 1 - Örnek

'dizi1' isimli bir DNA dizisi.

AGCTTTTCATTCTGACTGCAACGGGCAA
TATGTCTCTGTGTGGATTAAAAAAGAG
TGTCTGATAGCAGC

Problem 1 - Örnek

'dizi1' isimli bir DNA dizisi.

AGCTTTTCATTCTGACTGCAACGGGCAA
TATGTCTCTGTGTGGATTAAAAAAGAG
TGTCTGATAGCAGC

4 adet sayı: dizi1 içerisinde sırasıyla 'A',
'C', 'G' ve 'T' nükleotitlerinin görülme
sıklığı değerleri

A: 0.2857142857142857

C: 0.17142857142857143

G: 0.24285714285714285

T: 0.3

Problem 1

AGCTTTTTCATTCTGAAGCTTTTTCATTCTtggttgATTCTGAAGCATTCTGCCATT
CTGAAGCATTCTCATAGAACCCCTTCTCATAGAACCCCCGActccccctCTGTTC
TGAGAAAAAGACATCATTCTGTTCTGAGAAAAAGACATCATTCTGTTCTGAGA
AACCAAGACAaaaaggAGCTTTTTCATTCTGAAGCTTTTCATTCTGACaaggggagG
AAAAAGACATCATTCTGCATTCTGGGGAAAAGCTTTgggcatctttcaaCTGAAG
GCCATTCTGAACAAAAGACATCccat

Verilen DNA dizisi içerisindeki sırasıyla 'A', 'C', 'G' ve 'T' nükleotitlerinin görülme sıklığı değerlerini hesaplayınız.

Problem 2 - GC İçeriği (GC content)

$$\frac{G + C}{A + T + G + C} \times 100$$

Problem 2 - Örnek

'dizi2' isimli bir DNA dizisi.

AGCTTTTCATTCTGACTGCAACGGGCAA
TATGTCTCTGTGTGGATTAAAAAAGAG
TGTCTGATAGCAGC

Problem 2 - Örnek

'dizi2' isimli bir DNA dizisi.

AGCTTTTCATTCTGACTGCAACGGGCAA
TATGTCTCTGTGTGGATTAAAAAAGAG
TGTCTGATAGCAGC

3 adet sayı: dizi2 içerisindeki yüzde 'AT' içeriği değeri, dizi2 içerisindeki yüzde 'GC' içeriği değeri ve bulunan bu ilk iki değer toplamı.

AT içeriği: 58.57142857142858

GC içeriği: 41.42857142857143

Toplam: 100.0

Problem 2

1- GenBank ID'si **MG944891.1** olan DNA dizisini **NCBI veritabanı** üzerinden bulup, **FASTA** formatındaki verisine ulaşınız.

Problem 2

- 1- GenBank ID'si **MG944891.1** olan DNA dizisini **NCBI veritabanı** üzerinden bulup, **FASTA** formatındaki verisine ulaşınız.
- 2- Veriden kopyaladığınız DNA dizisini **Noccaea_papillosa** isimli bir değişkene atayınız.

Problem 2

- 1- GenBank ID'si **MG944891.1** olan DNA dizisini **NCBI veritabanı** üzerinden bulup, **FASTA** formatındaki verisine ulaşınız.
- 2- Veriden kopyaladığınız DNA dizisini **Noccaea_papillosa** isimli bir değişkene atayınız.
- 3- Bu DNA dizisinin yüzde AT içeriğini, yüzde GC içeriğini ve bu iki değer toplamını hesaplayınız.

Problem 3 - Tamamlayıcı DNA (Complementary DNA)

Main strand	ATGATTGACATTGAGGATCCAT
Ana Zincir	
Complementary Strand	TACTAACTGTAACCTAGGTA
Tamamlayıcı Zincir	

Sample genetic code with complementary strands.

© G.Osuri

Problem 3 - Örnek

'dizi3' isimli bir DNA dizisi.

ACTGATCGATTACGTATAGTATTTGCTATCATACATATATATCGATGCGTTCAT

Problem 3 - Örnek

'dizi3' isimli bir DNA dizisi.

ACTGATCGATTACGTATAGTATTTGCTATCATACATATATATCGATGCGTTCAT
TGACTAGCTAATGCATATCATAAACGATAGTATGTATATATAGCTACGCAAGTA

String; dizi3'ün tamamlayıcı zincirini içeren DNA dizisi.

Problem 3

1- GenBank ID'si **MG925459.1** olan DNA dizisini **NCBI veritabanı** üzerinden bulup, **FASTA** formatındaki verisine ulaşınız.

Problem 3

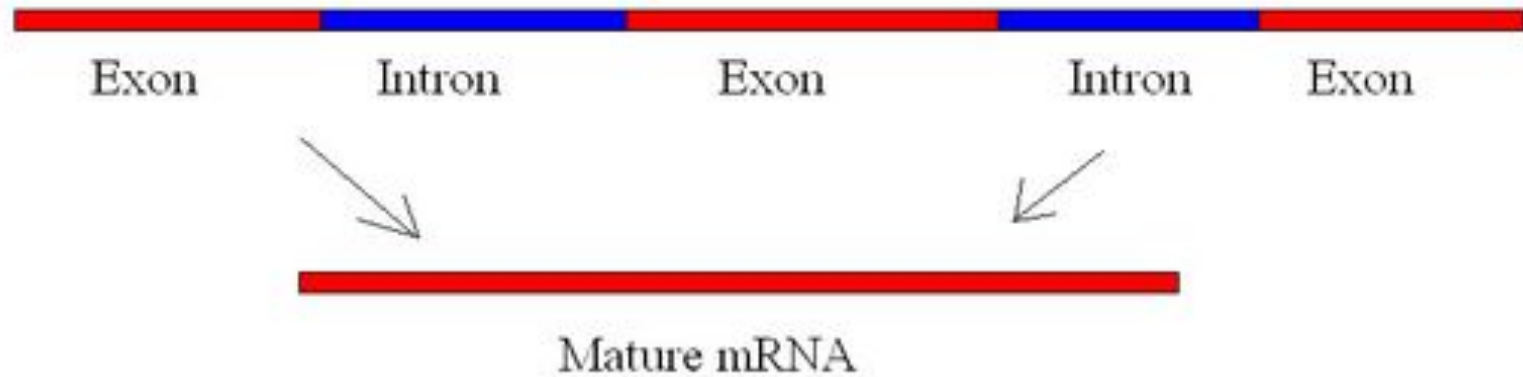
1- GenBank ID'si **MG925459.1** olan DNA dizisini **NCBI veritabanı** üzerinden bulup, **FASTA** formatındaki verisine ulaşınız.

2- Veriden kopyaladığınız DNA dizisini **Noccaea_papillosa_2** isimli bir değişkene atayınız.

Problem 3

- 1- GenBank ID'si **MG925459.1** olan DNA dizisini **NCBI veritabanı** üzerinden bulup, **FASTA** formatındaki verisine ulaşınız.
- 2- Veriden kopyaladığınız DNA dizisini **Noccaea_papillosa_2** isimli bir değişkene atayınız.
- 3- Bu DNA dizisinin tamamlayıcı zincirini bulunuz.

Problem 4 - Eksonlar ve intronlar



Problem 4 - Kısım 1: İntron'ları kesip atmak

Ekson ve intron bölgelerinden oluşan bir DNA dizisi verilmiştir.

ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCG
ATCGATCGATCGATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGA
CTACTAT

Problem 4 - Kısım 1: İntron'ları kesip atmak

Ekson ve intron bölgelerinden oluşan bir DNA dizisi verilmiştir.

ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCG
ATCGATCGATCGATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGA
CTACTAT

-İlk ekson dizinin başından 63. karaktere kadar olan kısmı kapsamaktadır.

Problem 4 - Kısım 1: İntron'ları kesip atmak

Ekson ve intron bölgelerinden oluşan bir DNA dizisi verilmiştir.

ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCG
ATCGATCGATCGATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGA
CTACTAT

-İlk ekson dizinin başından 63. karaktere kadar olan kısmı kapsamaktadır.

-İkinci ekson ise 91. karakterden dizinin sonuna kadar olan kısmı kapsamaktadır.

Problem 4 - Kısım 1: İntron'ları kesip atmak

Ekson ve intron bölgelerinden oluşan bir DNA dizisi verilmiştir.

```
ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCG  
ATCGATCGATCGATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGA  
CTACTAT
```

-İlk ekson dizinin başından 63. karaktere kadar olan kısmı kapsamaktadır.

-İkinci ekson ise 91. karakterden dizinin sonuna kadar olan kısmı kapsamaktadır.

Sadece kodlayan dizileri yazdıracak bir program yazınız.

Problem 4 - Kısım 2:

Kodlayan DNA yüzdesinin hesaplanması

Ekson ve intron bölgelerinden oluşan bir DNA dizisi verilmiştir.

ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCG
ATCGATCGATCGATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGA
CTACTAT

Birinci kısımda elde ettiğiniz veriyi kullanarak, verilen DNA dizisinin yüzde kaçının kodlayan dizi olduğunu hesaplayınız.

Problem 4 - Kısım 3:

DNA dizisini yeniden yazmak

Ekson ve intron bölgelerinden oluşan bir DNA dizisi verilmiştir.

ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCG
ATCGATCGATCGATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGA
CTACTAT

Birinci kısımda elde ettiğiniz veriyi kullanarak, orijinal DNA dizisini kodlayan bölgeleri büyük harf, kodlamayan bölgeleri küçük harf olacak şekilde tekrar yazdırınız.

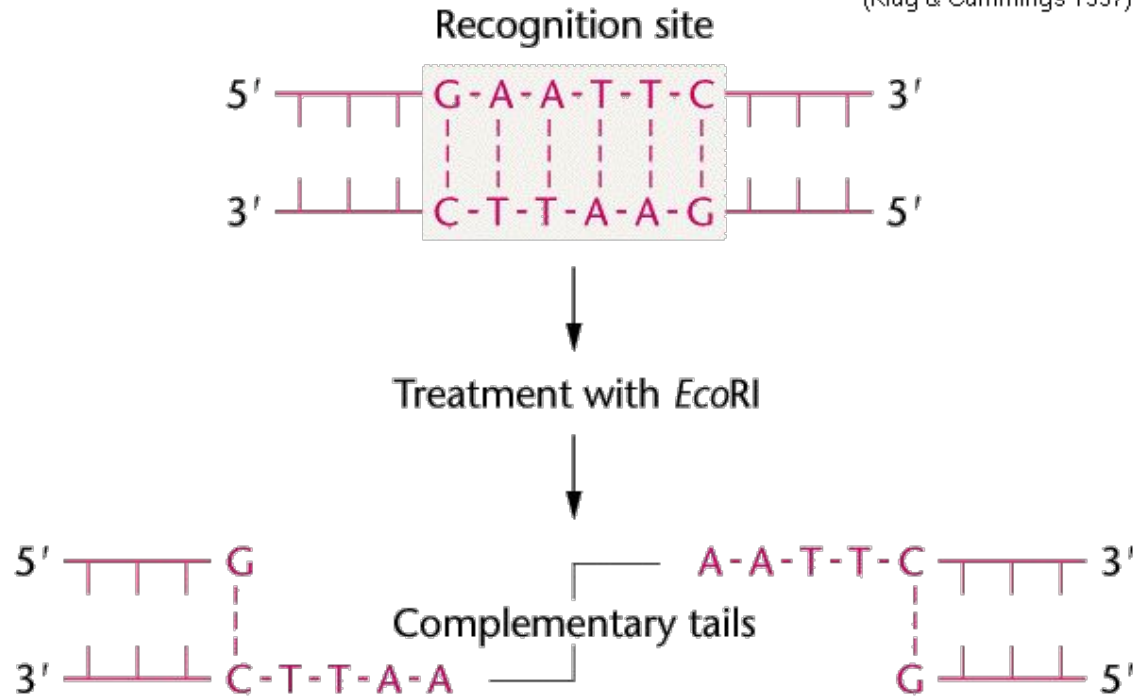
Problem 5

EcoRI bir kesim enzimidir. Bu kesim enzimi DNA'yı G*AATTC motifinin bulunduğu yerden keser. (* kesim yerini ifade etmektedir.)



Problem 5

(Klug & Cummings 1997)



Problem 5 - Örnek

Bir DNA dizisi.

GAATTC

Problem 5 - Örnek

Bir DNA dizisi.

GAATTC

Kesim sonucunda birinci parçanın uzunluğu 1 ve ikinci parçanın uzunluğu 5 olarak hesaplanmıştır.

Problem 5

ACTGATCGATTACGTATAGTAGAATTCTATCATACATATATATCGATGCGTTCAT

Verilen DNA dizisi EcoRI enzimi tarafından kesildikten sonra elde edilecek iki parçanın uzunluklarını hesaplayınız.

Problem 6 - Bonus Soru: Translasyon

Gerekli bilgiler:

- Dictionary
- Range fonksiyonu
- Döngüler (for loop, while loop)

İletişim: isinaltinkaya@gmail.com

Kaynaklar: <http://rsgturkey.com/en/resources/>