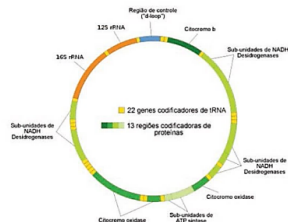


Análise de *codon usage bias* mitocondrial

Codon usage bias

- Código genético - Degenerado
- Múltiplos códons - Um mesmo aa
 - Último nucleotídeo - *Wobble*
- Preferência por códons específicos
 - Viés de uso de codon (*codon usage bias*)
 - Comum no genoma mitocondrial
- Possível explicação
 - Maximizar a transcrição de seus genes (e.g. mais A-T)



- Clado escolhido: Primates (~ 300 spp.)
 - Dados baixados pelo Portal RefSeq de organelas do NCBI
 - 50 mitogenomas escolhidos para a análise
- Análise dos códons que codificam para o aminoácido Tirosina
 - Dicotônico: TAT e TAC
 - Esperado: Tivesse uma maior porcentagem de TAT

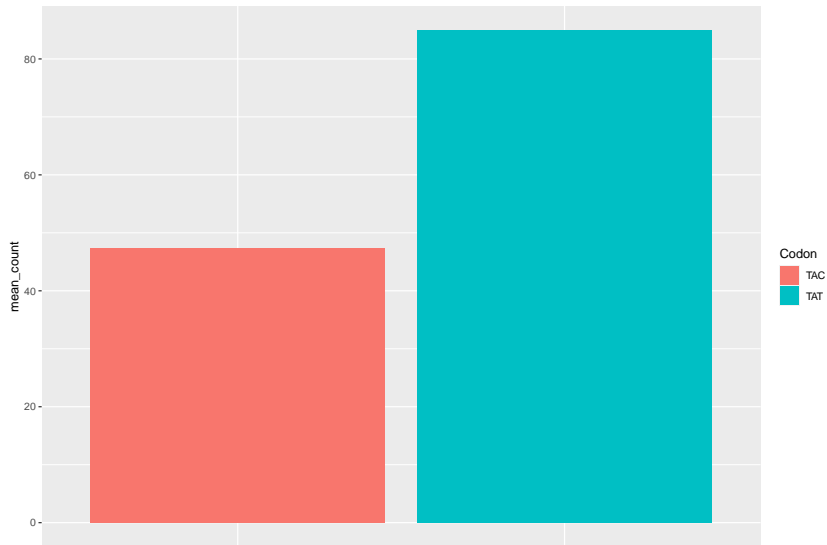
Tabela

● Primeiras linhas:

##	Species	Aminoacid	Codon	codon_count	/1000	Fraction
## 1	Cephalopachus_bancanus	Tyr	TAT	81	21.34	0.63
## 2	Cephalopachus_bancanus	Tyr	TAC	47	12.38	0.37
## 3	Cheirogaleus_crossleyi	Tyr	TAT	74	19.47	0.59
## 4	Cheirogaleus_crossleyi	Tyr	TAC	52	13.68	0.41
## 5	Cheirogaleus_medius	Tyr	TAT	81	21.28	0.60
## 6	Cheirogaleus_medius	Tyr	TAC	54	14.18	0.40
## 7	Eulemur_macaco_macaco	Tyr	TAT	86	22.64	0.66
## 8	Eulemur_macaco_macaco	Tyr	TAC	45	11.85	0.34
## 9	Eulemur_rufus	Tyr	TAT	75	19.75	0.58
## 10	Eulemur_rufus	Tyr	TAC	54	14.22	0.42

Visualização

```
## Codon mean_count mean_perc
## 1 TAC 47.38 0.3584
## 2 TAT 84.90 0.6416
```



- Qui-quadrado de aderência (*chi-square goodness of fit test*)
 - Limiar de significância: 0.05

```
chi_sqr_results = chisq.test(codon_summary$mean_count, p = c(0.5, 0.5))
```

```
chi_sqr_results
```

```
##
```

```
## Chi-squared test for given probabilities
```

```
##
```

```
## data: codon_summary$mean_count
```

```
## X-squared = 10.642, df = 1, p-value = 0.001105
```

- $p < 0.05$

Para a tirosina...

- Codons só variam na terceira posição
 - Esperaríamos que a escolha fosse aleatória
- Maximização da transcrição
 - Priorização do códon com mais AT

- E isso é tudo por enquanto

Resultado “real”

```
## Codon mean_count mean_perc
## 1 TAC 67.22111 0.5139698
## 2 TAT 63.78392 0.4859799
```



O p valor também não é o que vimos antes

```
chi_sqr = chisq.test(Tyr$mean_count,  
                     p = c(0.5, 0.5))  
chi_sqr  
##  
##  Chi-squared test for given probabilities  
##  
## data:  Tyr$mean_count  
## X-squared = 0.090182, df = 1, p-value = 0.7639
```

O que gerou o resultado significativo?

O que gerou o resultado significativo?

- A amostragem

O que gerou o resultado significativo?

- A amostragem
- A tabela original tem 199 espécies
- Eu escolhi as primeiras 50 espécies da tabela
- Mas isso não parece um problema
 - A menos que a ordem alfabética do nome das espécies esteja correlacionada com a preferência por codons TAT
 - Será que alguns gêneros possuem preferência por codons TAT?

O que gerou o resultado significativo?

- A amostragem
- A tabela original tem 199 espécies
- Eu escolhi as primeiras 50 espécies da tabela
- Mas isso não parece um problema
 - A menos que a ordem alfabética do nome das espécies esteja correlacionada com a preferência por codons TAT
 - Será que alguns gêneros possuem preferência por codons TAT?
- **Ou talvez eu tenha mexido na tabela antes...**

Manipulação dos dados

- Ordenei o dataset (os maiores valores de TAT primeiro)
- Selecionei as primeiras 50 espécies do dataframe

```
#Ordenando pelas espécies que possuem maior porcentagem de TAT
```

```
Tyr = dados %>%  
  filter(Aminoacid == "Tyr") %>%  
  arrange(desc(Codon), -Fraction)
```

```
#Extraindo as primeiras 50 espécies (que têm maior porcentagem TAT)
```

```
max_perc_tat = Tyr %>%  
  filter(Species %in% Tyr$Species[1:50]) %>%  
  group_by(Codon) %>%  
  summarise(mean_count = mean(codon_count), mean_perc = mean(Fraction))  
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
#Rodando o qui-quadrado
```

```
chi_sqr_sample = chisq.test(max_perc_tat$mean_count, p = c(0.5, 0.5))
```

```
chi_sqr_sample
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: max_perc_tat$mean_count  
## X-squared = 10.642, df = 1, p-value = 0.001105
```

The Nine Circles of Scientific Hell

- Forçei a barra (e não sei se deveria)
- Não sei ao certo qual pecado seria esse
 - Oitavo círculo: Publicação parcial
 - Nono círculo: Inventar dados
- Então, eu provavelmente ($p < 0.05$) estaria bem fundo no inferno científico