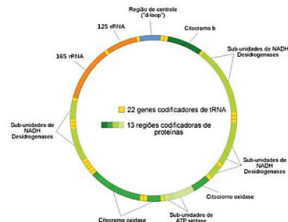


Análise de *codon usage bias* mitocondrial

Introdução

- Código genético - Degenerado
- Preferência por códons específicos
 - Viés de uso de codon (*codon usage bias*)
 - Comum no genoma mitocondrial
- Possíveis explicação
 - Maximizar a transcrição de seus genes (e.g. mais A-T)
 - Priorizar os códons do mitogenoma (22 tRNAs)
- Clado escolhido: Primates



Perguntas

Perguntas Biológicas

- 1 Quais aminoácidos apresentam *codon usage bias*?
- 2 Dentre os que apresentam:
 - a) Essa preferência está associada ao número de bases AT do códon?
 - b) Essa preferência está associada a quais codons o mitogenoma usa?

Perguntas experimentais

- 1 Há uma diferença entre o número de codons sinônimos usados para cada aa?
- 2
 - a) Há uma diferença entre o número de codons que possuem 1, 2 ou 3 Adeninas ou Timinas?
 - b) Há uma diferença entre o número de codons cujos tRNAs/anticodons são encontrados no mitogenoma e os que não são?

Obs: As perguntas são restritas à mitocôndrias de primatas.

Desenho - Estudo observacional

```
##                               Species Translation Table Aminoacid Codon Anticodon
## 1 Allenopithecus_nigroviridis           2      Phe   TTT      <NA>
## 2 Allenopithecus_nigroviridis           2      Phe   TTC       GAA
## 3 Allenopithecus_nigroviridis           2      Leu   TTA       TAA
## 4 Allenopithecus_nigroviridis           2      Leu   TTG      <NA>
## 5 Allenopithecus_nigroviridis           2      Leu   CTT      <NA>
## 6 Allenopithecus_nigroviridis           2      Leu   CTC      <NA>
##  codon_count /1000 Fraction
## 1           99 25.98      0.45
## 2          120 31.50      0.55
## 3          102 26.77      0.16
## 4           17  4.46      0.03
## 5           97 25.46      0.15
## 6          103 27.03      0.16
```

- Não usaremos todas as variáveis medidas
 - Variáveis Dependentes:
 - Codon count, Fraction
 - Variáveis Independentes:
 - Species, aminoacid, codon
- Unidade experimental: Espécie
 - Mitogenoma completo da espécie
 - Toda espécie consegue contribuir com um ponto por grupo - Critério de pareamento

Obtenção dos dados

- Clado escolhido: Primates
- Script python mitodownloader.py:
 - Baixa os dados pelo Portal RefSeq de organelas do NCBI
- tRNAscan-SE usado para identificar os anticodons
- Python de novo:
 - Pandas e Biopython
 - Calcular métricas de codon usage
 - Exportar para uma **tabela**
- Tudo disponível em:

https://github.com/gavieira/projeto_bioestatistica

Tabela

- Primeiras 6 linhas:

```
##               Species Translation Table Aminoacid Codon Anticodon
## 1 Allenopithecus_nigroviridis          2      Phe   TTT      <NA>
## 2 Allenopithecus_nigroviridis          2      Phe   TTC      GAA
## 3 Allenopithecus_nigroviridis          2      Leu   TTA      TAA
## 4 Allenopithecus_nigroviridis          2      Leu   TTG      <NA>
## 5 Allenopithecus_nigroviridis          2      Leu   CTT      <NA>
## 6 Allenopithecus_nigroviridis          2      Leu   CTC      <NA>
##   Number of Codon Occurences /1000 Fraction
## 1                99 25.98      0.45
## 2               120 31.50      0.55
## 3               102 26.77      0.16
## 4                17  4.46      0.03
## 5                97 25.46      0.15
## 6               103 27.03      0.16
```

- Dimensões da tabela:

```
## [1] 12736      8
```

- Número de espécies

```
## [1] 199
```

O que fazer com isso?

Eu também não sei direito.

Hipótese 1 - Mitocôndria prioriza codons complementares aos seus tRNAs

Comparar os códons complementares aos tRNAs mitocondriais com os outros

Hipótese 2 - Códon com mais A-T são priorizados

Comparar os códons com 1, 2 ou 3 posições que possuam adenina ou timina entre si

Mas ainda falta ler mais sobre o assunto e (claro) aprender mais sobre estatística. . .