

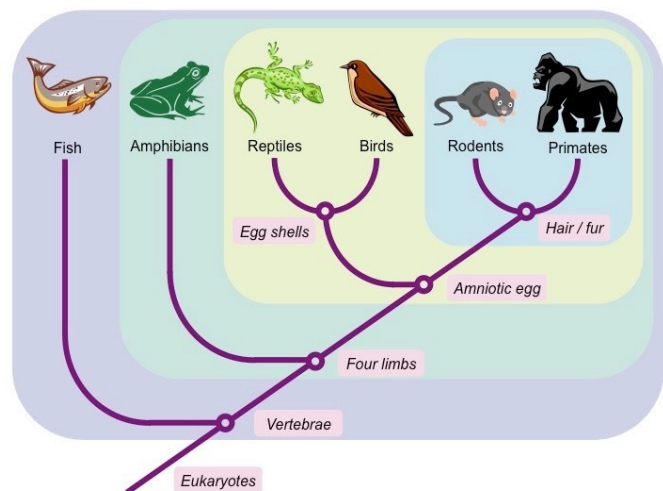
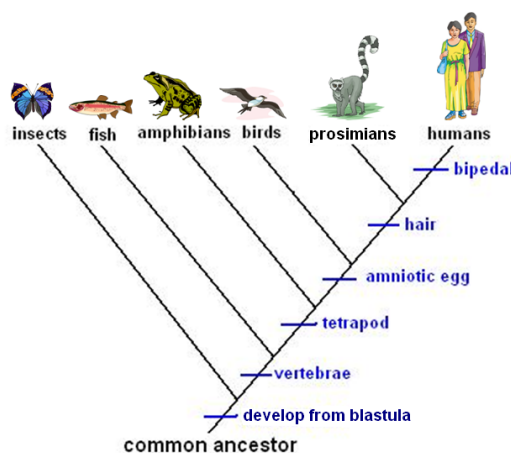
MC102QR - Algoritmos e Programação de Computadores

Lab 10 - Combinando Estruturas de Dados

Prazo da atividade: 12 de Junho de 2022

Peso na nota: 3 (7,32%)

Neste laboratório você utilizará seus conhecimentos de programação para ajudar em uma das atividades mais interessantes da biologia, que é a construção de diagramas de filogenia. O diagrama mais comum de filogenia é o Cladograma¹ (às vezes chamado de árvore de ancestralidade), nele é visto em forma de árvore a relação de ancestralidade entre grupos de animais. Abaixo alguns exemplos de cladogramas:



Como pode ser visto nos diagramas, os grupos, ou espécies, são agrupados a partir de características mais antigas onde prováveis ancestrais em comum introduziram na evolução. Por exemplo, na imagem da direita o que separa os peixes de todos os grupos é o fato de não possuírem quatro membros; já o que separa roedores e primatas é possuírem pêlo. Atualmente a melhor forma de construir cladogramas é por comparação de sequência de DNA, gerando vários problemas computacionais da chamada filogenética computacional².

Você foi designado a fazer o cladograma a partir de um conjunto de amostras de DNA, e quais características elas codificam, de várias espécies alienígenas que foram roubadas da área 51 e divulgadas ao público. Essas amostras contém informações sobre as espécies, mas saber a

¹ <https://pt.wikipedia.org/wiki/Cladograma>

² https://pt.wikipedia.org/wiki/Filogen%C3%A9tica_computacional

ancestralidade de cada uma ainda é um desafio. Por sorte se sabia das amostras qual o DNA mais antigo e o seu trabalho é contar a quantidade de mutações ocorridas desde do DNA ancestral. Quanto mais mutações, mais recente é considerado o DNA. Uma propriedade interessante do DNA alienígena (DNAA) é que ele é helicoidal, ou seja, formado por pares de proteínas que estão pareadas em formato de dupla hélice. Além das amostras, várias fichas com informações básicas foram encontradas e com elas você pode agrupar as informações do alienígena de acordo com suas características do cladograma. Com isso você poderá mostrar ao mundo que a verdade estava lá fora esse tempo todo³.

Tarefa

Seu programa deve agrupar as espécies de acordo com a ancestralidade e depois imprimir a características mais ancestrais e as espécies que são definidas por essas características. Portanto a tarefa ficaria em encontrar a ordem das características, contando o número de mutações em relação ao DNAA mais antigo, após isso, identificar quais espécies são caracterizadas por qual característica e imprimir por ordem decrescente de ancestralidade. Pegando o exemplo da esquerda na imagem de exemplo temos:

- Desenvolvimento por blástulas: Insetos
- Vertebrado: Peixes
- Tetrapode: Anfíbios
- Ovos amnióticos: Aves
- Pêlos: Simios
- Bípede: Humanos

A diferença é que você irá contar mutações de DNAA e existem alguns aspectos do DNAA que você precisa entender para fazer isso:

1. O DNAA é formado por pares de proteínas base que se ligam.
2. Diferente dos humanos, são 5 proteínas que formam o DNAA com as letras "B", "C", "D", "E", "F"; além disso cada base se liga com qualquer outra base.
3. As mutações para o DNAA são contadas levando não somente a diferença mas também a ordem do par, onde pares menores (lexicograficamente) não são considerados mutações. Por exemplo, o par "CB" não é considerado mutação do par "CF", mas "DE" é considerado mutação de "CF".

Levando isso em consideração, é possível contar quais características têm mais mutações em relação à característica ancestral e, portanto, são mais recentes. Não haverá caso de características com a mesma quantidade de mutações. Por fim, as espécies podem ser agrupadas pela diferença de suas características e ordenada pelas características mais antigas.

Entrada

A entrada para a formação do cladograma será formada pelos seguintes componentes:

- Quantidade de espécies a serem utilizadas no cladograma.

³ Essa frase é apenas para propósitos narrativos. Não acredite em teorias da conspiração.

- Lista de espécies e suas características. Os ids são identificadores numéricos. Formato:
 - `id nome_esécie nome_científico "|" carac_1, carac_2, ..., carac_n`
- Quantidade de características a serem levadas em conta.
- Nome da característica e duas sequências de letras (de mesmo tamanho) representando cada fita de DNA na linha seguinte. Ou seja, um par seria formado por uma letra da primeira linha e a outra letra na mesma posição da linha seguinte. A característica mais ancestral é sempre a primeira na entrada, depois desta não há ordem de ancestralidade garantida. Exemplo:

```
carac_1
BFEDCBFEDBBDDF
FDCBDFDCBFDDBB
```

Saída

A saída do programa é a estrutura de grupos do cladograma. Isso significa o nome da característica e o nome das espécies que contém exclusivamente essa característica como mais ancestral. Os grupos devem ser ordenados por ordem decrescente de ancestralidade. As espécies são ordenadas ascendentemente por id. Assim a saída deve ser no formato:

```
CARACTERÍSTICA: X
id1 nome_esécie1 nome_científico1
id2 nome_esécie2 nome_científico2
...
=====
CARACTERÍSTICA: Y
id3 nome_esécie3 nome_científico3
id4 nome_esécie4 nome_científico4
...
=====
```

Neste exemplo X é uma característica mais ancestral do que Y. Tal estrutura deve seguir para todas as características citadas na entrada.

Observações:

- A partir do Python 3.7 a ordem de inserção em dicionário é garantida, ou seja, você pode iterar sobre um dicionário seguindo a ordem de inserção das chaves. Previamente isso era feito utilizando `OrderedDict` da biblioteca `collections`.
- Uma das formas de ordenar listas é utilizar a função `sorted`⁴ que recebe de entrada uma lista e retorna a lista ordenada.
Ex: `sorted([5, 2, 3, 1, 4]) -> [1, 2, 3, 4, 5]`

⁴ <https://docs.python.org/pt-br/dev/howto/sorting.html>

Exemplos

Exemplo 1:

Entrada

```
3
143 Xenomorfo Xenomorfesium | ovos_parasitas bipede
125 Pleiadian Pleiademus | bipede
176 Asteromorfo Asteromorfesium | hiper-inteligencia
3
bipede
BFEDCBFEDBBDDFC
FDCBDFDCBFDCCBB
ovos_parasitas
BFEECBFEDBBDDFB
FDFCDFDCBFDCCBC
hiper-inteligencia
BFFFECFEDBBDDFC
FDCBDBCBFDCBB
```

Saída

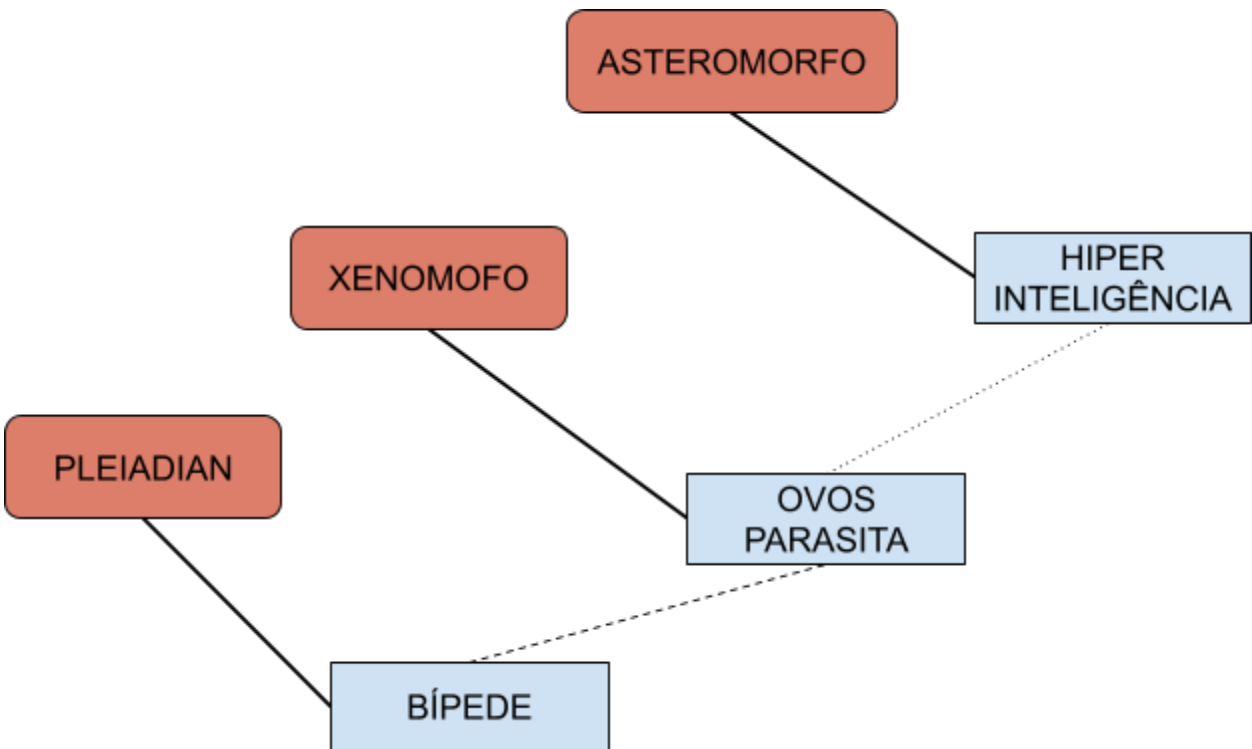
```
CARACTERÍSTICA: bipede
125 Pleiadian Pleiademus
CARACTERÍSTICA: ovos_parasitas
143 Xenomorfo Xenomorfesium
CARACTERÍSTICA: hiper-inteligencia
176 Asteromorfo Asteromorfesium
```

Passa a Passo Exemplo 1:

1. Processar entrada.
2. Contar as mutações tendo como base a característica mais antiga, que é sempre a primeira, no caso a característica `bipede`. Neste exemplo, os pares (E,F) e (E,C) da característica `ovos_parasitas` e os pares (F, C), (F, B), (E, D) e (C, B) em `hiper-inteligencia`. Marcados de amarelo no texto do exemplo.
3. Dado a quantidade de mutações, temos a ordem da ancestralidade (maior quantidade de mutações é mais recente). Logo a ordem `bipede`, `ovos_parasitas` e `hiper-inteligencia`.
4. Dados esta ordem, identificar quais características diferenciam unicamente uma espécie alienígena seguindo a ordem de ancestralidade. No exemplo: `Pleiadian` tem como característica apenas `bipede` logo está neste grupo. `Xenomorfo` tem a característica

bidepe e ovos_parasitas mas como a característica mais recente é ovos_parasitas, apareceu depois na ordem de características, é esta que define o grupo do Xenomorfo. Por fim Asteromorfo contém a característica hiper-inteligencia e esta é a característica mais recente então é agrupado por ela.

5. Imprimir na tela seguindo o padrão especificado na seção Saída.



Representação em árvore do exemplo 1.

Exemplo 2:

Entrada

```

5
143 Xenomorfo Xenomorfesium | ovos_parasitas bipede
122 Pleiadian Pleiademus | bipede
543 Asteromorfo Asteromorfesium | hiper-inteligencia
132 Eosapiens Eosapienses | pele_azul hiper-inteligencia
565 Malakak Malakamium | bipede pele_azul hiper-inteligencia
4
bipede
BFEDCBFEDBBDDFC
FDCBDFDCBFDCBBD
ovos_parasitas
  
```

```
BFEECBFEDBBDDFC
FDFCDFDCBFDCBBB
hiper-inteligencia
BFFFECFEDBBDDFC
FDCBDBBCBFDCBBB
pele_azul
BFFFECFEDBBDDFC
FDCBDBBCBFDCBBF
```

Saída

```
CARACTERÍSTICA: bipede
122 Pleiadian Pleiademus
CARACTERÍSTICA: ovos_parasitas
143 Xenomorfo Xenomorfesium
CARACTERÍSTICA: hiper-inteligencia
543 Asteromorfo Asteromorfesium
CARACTERÍSTICA: pele_azul
132 Eosapiens Eosapienses
565 Malakak Malakamium
```

Exemplo 3:

Entrada

```
5
143 Xenomorfo Xenomorfesium | ovos_parasitas bipede
122 Pleiadian Pleiademus | bipede
543 Asteromorfo Asteromorfesium | hiper-inteligencia
132 Eosapiens Eosapienses | pele_azul hiper-inteligencia
asas_tentaculos
565 Malakak Malakamium | bipede pele_azul hiper-inteligencia
5
bipede
BFEDCBFEDBBDDFC
FDCBDFDCBFDDBBB
ovos_parasitas
BFEECBFEDBBDDFC
FDFCDFDCBFDDBBB
hiper-inteligencia
BFFFECFEDBBDDFC
FDCBDBBCBFDDBBB
pele_azul
BFFFECFEDBBDDFC
FDCBDBBCBFDDBBF
asas_tentaculos
CFEFECDEDBDDFC
FBDBDBBCBFDDBBF
```

Saída

```
CARACTERÍSTICA: bipede
122 Pleiadian Pleiademus
CARACTERÍSTICA: ovos_parasitas
143 Xenomorfo Xenomorfesium
CARACTERÍSTICA: hiper-inteligencia
543 Asteromorfo Asteromorfesium
CARACTERÍSTICA: pele_azul
565 Malakak Malakamium
CARACTERÍSTICA: asas_tentaculos
132 Eosapiens Eosapienses
```

Submissão

Você deverá submeter no CodePost, na tarefa Lab 10, um arquivo com o nome `lab10.py`, contendo todo o seu programa.

Após o prazo estabelecido para a atividade, será aberta uma tarefa Lab 10 - Segunda Chance, com prazo de entrega até o fim do semestre.