

ENSINO MÉDIO INTEGRADO – INFORMÁTICA

LINGUAGEM DE PROGRAMAÇÃO

*Manipulando Arquivos Texto
(Trabalho 10)*



BUSCA GENÉTICA

- *A atividade proposta utiliza-se de informações totalmente hipotéticas, e que não possuem valor científico dentro da biologia. O único intuito do trabalho é colocar o aluno perante em uma situação onde necessite utilizar os conceitos abordados sobre arquivos texto dentro da linguagem C.*

1. Introdução – Contextualização

Um laboratório está efetuando estudos sobre a sequência genética de três novas espécies de aranhas que foram descobertas, mas ainda não foram catalogadas. Essas aranhas foram encontradas em diferentes regiões do planeta e foram batizadas como:

- ✓ **Aranha Albina:** recebeu esse nome por não apresentar pigmentação (cor branca). Foi encontrada no interior de cavernas escuras no Oriente Médio, onde havia pouca ou nenhuma luminosidade.
- ✓ **Aranha Escorpião:** recebeu esse nome por possuir um prolongamento em seu exoesqueleto que lembra a cauda de um escorpião. Foi encontrada na América do Sul em regiões de clima quente.
- ✓ **Aranha Cigarra:** recebeu esse nome por emitir um som agudo, parecido com o de uma cigarra, com objetivo de atordoar sua presa antes de efetuar o ataque. Foi encontrada na região central da África em local que possui alta umidade relativa do ar.

Algumas informações já foram pré-determinadas, considerando que se tratam de três aranhas, elas possuem as seguintes classificações:

- **Reino:** Animalia
- **Filo:** Arthropoda
- **Classe:** Arachnida
- **Ordem:** Araneae

Contudo, o laboratório que está efetuando a análise genética das três novas espécies precisa determinar em quais famílias e gêneros elas se enquadram.

2. Atividade a ser desenvolvida

Os biólogos que chefiam o laboratório já descobriram, com análises preliminares, que as aranhas estudadas pertencem a três possíveis famílias e três possíveis gêneros, de acordo com os seguintes padrões genéticos:

Família

- ❖ ***Theridiidae***: o número de sequências AACGTCT > TCGACGG e CCTTGAA
- ❖ ***Ctenidae***: o número de sequências TCGACGG > AACGTCT e CCTTGAA
- ❖ ***Theraphosidae***: o número de sequências CCTTGAA > AACGTCT e TCGACGG

Gênero

- ❖ ***Latrodectus***: o número de sequências TTAACGC > GGATACC e CGCGTTA
- ❖ ***Phoneutria***: o número de sequências GGATACC > TTAACGC e CGCGTTA
- ❖ ***Avicularia***: o número de sequências CGCGTTA > TTAACGC e GGATACC

O mapeamento genético das espécies já foi efetuado e armazenado em três arquivos texto. Para segunda etapa do trabalho os biólogos responsáveis pela pesquisa estão contratando você para desenvolver um aplicativo que efetue a leitura dessas três sequências e ao final gere um novo arquivo texto, denominado “catalogo.txt”, que conterà os nomes das novas espécies e a qual família e gênero elas pertencem. O arquivo final deve conter um formato exatamente igual ao mostrado a seguir:

ARANHA ALBINA

(família / gênero)

[Família]

AACGTCT: X vezes

TCGACGG: Y vezes

CCTTGAA: Z vezes

[Gênero]

TTAACGC: A vezes

GGATACC: B vezes

CGCGTTA: C vezes

ARANHA ESCORPIÃO

(família / gênero)

[Família]

TCGACGG: X vezes

AACGTCT: Y vezes

[Gênero]

GGATACC: A vezes

TTAACGC: B vezes

CCTTGAA: Z vezes

CGCGTTA: C vezes

ARANHA CIGARRA

(família / gênero)

[Família]

CCTTGAA: X vezes

TCGACGG: Y vezes

AACGTCT: Z vezes

[Gênero]

CGCGTTA: A vezes

TTAACGC: B vezes

GGATACC: C vezes

OBS.: onde aparece (família / gênero) deve ser substituído pelo nome da família e do gênero descoberto para que cada uma das espécies analisadas, por exemplo: (Ctenidae / Avicularia). Os caracteres X, Y, Z, A, B e C devem ser substituídos pelo número de vezes em que a sequência que eles representam apareceu dentro do código genético mapeado das três aranhas.

RESULTADOS – ARQUIVO TEXTO DE SAÍDA

```
1 ARANHA ALBINA
2 (Theridiidae / Phoneutria)
3
4 Sequencias - Familia          Sequencias - Genero
5 AACGTCT:      64 vezes        TTAACGC:      53 vezes
6 TCGACGG:      59 vezes        GGATACC:      68 vezes
7 CCTTGAA:      57 vezes        CGCGTTA:      62 vezes
8
9 ARANHA ESCORPIAO
10 (Theraphosidae / Latrodectus)
11
12 Sequencias - Familia          Sequencias - Genero
13 AACGTCT:      55 vezes        TTAACGC:      71 vezes
14 TCGACGG:      59 vezes        GGATACC:      56 vezes
15 CCTTGAA:      72 vezes        CGCGTTA:      50 vezes
16
17 ARANHA CIGARRA
18 (Theridiidae / Avicularia)
19
20 Sequencias - Familia          Sequencias - Genero
21 AACGTCT:      66 vezes        TTAACGC:      53 vezes
22 TCGACGG:      56 vezes        GGATACC:      62 vezes
23 CCTTGAA:      61 vezes        CGCGTTA:      64 vezes
```