

Notebook

September 22, 2025

1 Trabalho 1: Alinhamento de Sequências

Leonardo Azzi Martins

As sequências de entrada foram armazenadas no diretório **sequences**, onde os arquivos são nomeados como **<tax_id>.txt**. Os **tax_id** 01 e 02 são os exemplos de teste fornecidos pelo professor para toda a turma.

2 Questão 1.

- a) Desenvolva um programa que implemente o algoritmo de Needleman-Wunsch. A implementação deve incluir a criação da matriz de pontuação e a função de traceback para reconstruir o alinhamento.

Os programas estão em `global_alignment.py`.

- b) Utilize a função de alinhamento para processar os seguintes pares de sequências:

Homo Sapiens vs. Chrysocyon brachyurus Homo Sapiens vs. Gallus gallus Homo Sapiens vs. Oncorhynchus mykiss

- Humano (Homo sapiens) [TaxId: 9606]
- Lobo-Guará (Chrysocyon brachyurus) [TaxId: 68728]
- Galinha (Gallus gallus) [TaxId: 9031]
- Truta Arco-Íris (Oncorhynchus mykiss) [TaxId: 8022]

Para cada par, imprima o alinhamento global resultante e a pontuação final obtida.

```
[3]: import global_alignment

humano_vs_lobo_guara = global_alignment.main("9606", "68728")

print("Alinhamento global: Homo Sapiens vs. Chrysocyon brachyurus")
print("Sequência 1:", humano_vs_lobo_guara["seq_1"])
print("Sequência 2:", humano_vs_lobo_guara["seq_2"])
print("Pontuação final:", humano_vs_lobo_guara["score"], "\n")

humano_vs_galinha = global_alignment.main("9606", "9031")

print("Alinhamento global: Homo Sapiens vs. Gallus gallus")
```

```

print("Sequência 1:", humano_vs_galinha["seq_1"])
print("Sequência 2:", humano_vs_galinha["seq_2"])
print("Pontuação final:", humano_vs_galinha["score"], "\n")

humano_vs_truta = global_alignment.main("9606", "8022")

print("Alinhamento global: Homo Sapiens vs. Oncorhynchus mykiss")
print("Sequência 1:", humano_vs_truta["seq_1"])
print("Sequência 2:", humano_vs_truta["seq_2"])
print("Pontuação final:", humano_vs_truta["score"])

```

Alinhamento global: Homo Sapiens vs. Chrysocyon brachyurus

Sequência 1: VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALT
NAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY-

Sequência 2: VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPPTTKTYFPHFDLSPGSAQVKAHGKKVADALT
TAVAHLLDPLGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFTAVSTVLTSKYR

Pontuação final: 181

Alinhamento global: Homo Sapiens vs. Gallus gallus

Sequência 1: VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALT
NAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY-

Sequência 2: MLTAEDKKLIQAWWEKAASHQEEFGAEALTRMFTTYPQTKTYFPHFDLSPGSDQVRGHGKKVLGALG
NAVKNVDNLSQAMAELSNLHAYNLRVDPVNFKLLSQCIQVVLAVHMGKDYPTEVHAADFDFLSAVSAVLAEKYR

Pontuação final: 49

Alinhamento global: Homo Sapiens vs. Oncorhynchus mykiss

Sequência 1: -VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPH-FDLSHGSAQVKGHGKKVADA
LTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY-

Sequência 2: XSLTAKDKSVVKAFFWGKISGKADVVGAELGRMLTAYPQTKTYFSHWADLSPGSGPVKKHGGIIMGA
IGKAVGLMDDLVGMSALSDLHAFKLRVDPGNFKILSHNILLVTLAIHFPSDFTPEVHIAVDKFLAAVSAALADKYR

Pontuação final: 31

c) Calcule a porcentagem de identidade de sequências para cada um dos pares acima

```

[4]: print("Identidade: Homo Sapiens vs. Chrysocyon brachyurus")
print("Porcentagem de identidade:", f"{humano_vs_lobo_guara['identity']*100:.  
↪3f}%", "\n")

print("Identidade: Homo Sapiens vs. Gallus gallus")
print("Porcentagem de identidade:", f"{humano_vs_galinha['identity']*100:.  
↪3f}%", "\n")

print("Identidade: Homo Sapiens vs. Oncorhynchus mykiss")
print("Porcentagem de identidade:", f"{humano_vs_truta['identity']*100:.3f}%")

```

Identidade: Homo Sapiens vs. Chrysocyon brachyurus

Porcentagem de identidade: 82.270%

Identidade: Homo Sapiens vs. Gallus gallus

Porcentagem de identidade: 58.865%

Identidade: Homo Sapiens vs. Oncorhynchus mykiss

Porcentagem de identidade: 55.944%

- d) Baseado nos resultados, identifique qual espécie apresenta a maior similaridade com a sequência Homo Sapiens.

No alinhamento global, a espécie Chrysocyon brachyurus apresentou 181 pontos de alinhamento global, e 82,27% de identidade.

3 Questão 2

- Valor para GAP: -3
- Valor para Match: +2
- Valor para Mismatch: -2
- Prioridade no traceback deve ser a diagonal sempre que houver um empate
- Se houver mais de uma subsequência no algoritmo de Smith-Waterman com o mesmo score, retorne a primeira encontrada

- a) Implemente o algoritmo de Smith-Waterman para alinhamento local. Sua implementação deve incluir a criação da matriz de pontuação e a função de traceback.

Implementado em local_alignment.py

- b) Identifique e extraia a sub-sequência com a maior pontuação para cada par:

Homo Sapiens vs. Chrysocyon brachyurus Homo Sapiens vs. Gallus gallus Homo Sapiens vs. Oncorhynchus mykiss

Qual é o valor dessa pontuação e qual é a sub-sequência correspondente?

```
[5]: import local_alignment

humano_vs_lobo_guara = local_alignment.main("9606", "68728")

print("Alinhamento local: Homo Sapiens vs. Chrysocyon brachyurus")
print("Sequência 1:", humano_vs_lobo_guara["seq_1"])
print("Sequência 2:", humano_vs_lobo_guara["seq_2"])
print("Pontuação final:", humano_vs_lobo_guara["score"], "\n")

humano_vs_galinha = local_alignment.main("9606", "9031")

print("Alinhamento local: Homo Sapiens vs. Gallus gallus")
print("Sequência 1:", humano_vs_galinha["seq_1"])
print("Sequência 2:", humano_vs_galinha["seq_2"])
print("Pontuação final:", humano_vs_galinha["score"], "\n")

humano_vs_truta = local_alignment.main("9606", "8022")

print("Alinhamento local: Homo Sapiens vs. Oncorhynchus mykiss")
```

```
print("Sequência 1:", humano_vs_truta["seq_1"])
print("Sequência 2:", humano_vs_truta["seq_2"])
print("Pontuação final:", humano_vs_truta["score"])
```

Alinhamento local: Homo Sapiens vs. Chrysocyon brachyurus

Sequência 1: VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALT
NAVAHVDDMPNALSALSDDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY

Sequência 2: VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHFDLSPGSAQVKAHGKKVADALT
TAVAHLLDPLGALSALSDDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFTAVSTVLTSKY

Pontuação final: 184

Alinhamento local: Homo Sapiens vs. Gallus gallus

Sequência 1: GAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVL

Sequência 2: GAEALTRMFTTYPQTKTYFPHFDLSPGSDQVRGHGKKVLGALGNAVKNVDNLSQAMAELSNLHAYNL
RVDPVNFKLLSQCIQVVLAVHMGKDYTPEVHAAFDKFLSAVSAVL

Pontuação final: 68

Alinhamento local: Homo Sapiens vs. Oncorhynchus mykiss

Sequência 1: SALSDDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS

Sequência 2: SALSDDLHAFKLRVDPGNFKILSHNILVTLAIHFPSDFTPEVHIAVDKFLAAVS

Pontuação final: 50

- c) Realize o traceback a partir da célula na linha 4 e coluna 5 da matriz de pontuação. Quais os alinhamento local obtidos para cada par avaliado?

```
[6]: import local_alignment

start_pos = (4, 5) # Linha 4, coluna 5

humano_vs_lobo_guara = local_alignment.main("9606", "68728", start_pos)

print("Alinhamento local: Homo Sapiens vs. Chrysocyon brachyurus")
print("Sequência 1:", humano_vs_lobo_guara["seq_1"])
print("Sequência 2:", humano_vs_lobo_guara["seq_2"])
print("Pontuação final:", humano_vs_lobo_guara["score"], "\n")

humano_vs_galinha = local_alignment.main("9606", "9031", start_pos)

print("Alinhamento local: Homo Sapiens vs. Gallus gallus")
print("Sequência 1:", humano_vs_galinha["seq_1"])
print("Sequência 2:", humano_vs_galinha["seq_2"])
print("Pontuação final:", humano_vs_galinha["score"], "\n")

humano_vs_truta = local_alignment.main("9606", "8022", start_pos)

print("Alinhamento local: Homo Sapiens vs. Oncorhynchus mykiss")
print("Sequência 1:", humano_vs_truta["seq_1"])
```

```
print("Sequência 2:", humano_vs_truta["seq_2"])
print("Pontuação final:", humano_vs_truta["score"])
```

Alinhamento local: Homo Sapiens vs. Chrysocyon brachyurus

Sequência 1: VLSP-

Sequência 2: VLSPA

Pontuação final: 5

Alinhamento local: Homo Sapiens vs. Gallus gallus

Sequência 1:

Sequência 2:

Pontuação final: 0

Alinhamento local: Homo Sapiens vs. Oncorhynchus mykiss

Sequência 1:

Sequência 2:

Pontuação final: 0

4 Questão 3

- a) A partir das sequências alinhadas, qual conclusão podemos chegar ao analisar o score e a identidade obtida?

As sequências selecionadas para diferentes espécies são hemoglobinas (subunit alpha). Tanto com alinhamento global quanto local, o Lobo-guará apresentou uma sequência mais próxima à do Humano. Podemos considerar que esta proximidade tem explicação evolutiva/filogenética, pois ambas espécies são mamíferos, quanto o Galo e a Truta pertencem a espécies distintas.

- b) Qual a diferença do score do alinhamento e a identidade?

O score do alinhamento representa a soma dos valores reconstituídos no traceback, que são relativos a matches, mismatches e gaps ao longo do alinhamento de acordo com os valores de pontuação definidos. Já a identidade corresponde à proporção de matches entre as duas sequências alinhadas, expressa em porcentagem. O score acaba servindo como uma medida de similaridade ou “distância de edição” entre duas sequências, e a identidade é útil para avaliar o grau de conservação entre as sequências.

- c) Quais são as aplicações do alinhamento de sequências (seja em sequências de nucleotídeos ou de aminoácidos)? Quais seriam os possíveis passos seguintes de um alinhamento de sequência?

O alinhamento de sequências é utilizado na bioinformática para identificar regiões conservadas, inferir relações evolutivas, detectar mutações, prever funções de genes/proteínas e realizar anotação funcional. É utilizado em filogenia, análise de variantes, busca por homologia, montagem de genomas. entre outras aplicações. Após o alinhamento, os próximos passos podem incluir: análise filogenética, identificação de domínios conservados, anotação funcional, detecção de variantes, construção de árvores evolutivas, busca por regiões alvo para edição gênica, ou integração dos resultados em bancos de dados para estudos comparativos.

This notebook was converted with convert.ploomber.io