Trabalho 2: Filogenia

Leonardo Azzi Martins

a) Descubra o organismo correspondente de cada uma das 10 sequência (dica: usar o Blast para esta tarefa, https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Formato da Resposta:

seq_1: Streptococcus agalactiae

seq_2: Neisseria gonorrhoeae

seq_3: Mycobacterium tuberculosis

seq_4: Staphylococcus aureus

seq_5: Treponema pallidum

seq_6: Bacillus anthracis

seq_7: Yersinia pestis

seq_8: Acinetobacter baumannii

seq_9: Pseudomonas aeruginosa

seq_10: Helicobacter pylori

b) Realize o alinhamento global par-a-par de todas as sequências entre si. Para esta tarefa deve ser utilizado o algoritmo Needleman e Wunsch implementado no Trabalho I da disciplina. Você deve reportar o score de cada alinhamento considerando o seguinte esquema de pontuação: Match = +1; Mismatch = -1; Gap = -2.

Formato da Resposta:

seq_1 vs seq_2: xxx seq_1 vs seq_3: xxx ... seq_10 vs seq_9: xxx

In [1]:
```
import numpy as np
sequences_ids = list(range(1, 11))
matrix_size = len(sequences_ids)

distance_matrix = np.zeros((matrix_size, matrix_size))
```

In [2]:
```
# b)
import global_alignment
import numpy as np
```

```
dir = "./2_phylogenetics"
format = "fasta"
id1 = "seq_1"
id2 = "seq_2"
scoring_file = "scoring"

sequences_ids = list(range(1, 11))
num_of_seqs = len(sequences_ids)

distance_matrix = np.zeros((num_of_seqs, num_of_seqs))

for target_id in sequences_ids:
    print("Alvo:", target_id)

    candidates_ids = sequences_ids.copy()
    candidates_ids.remove(target_id)
    print("Candidatos:", candidates_ids)

    for candidate_id in candidates_ids:
        print(f"\nseq_{target_id} vs seq_{candidate_id}\n--------------------
        result = global_alignment.main(dir, format, f"seq_{target_id}", f"se
        print(result)

        # c)
        distance_matrix[target_id - 1, candidate_id - 1] = result["score"]
        distance_matrix[candidate_id - 1, target_id - 1] = result["score"]

    print("\n==========================================================
    distance_matrix
```

```
Alvo: 1
Candidatos: [2, 3, 4, 5, 6, 7, 8, 9, 10]

seq_1 vs seq_2
----------------------------------------------------------------------
{'seq_1': 'AAAC-ACCTCCAGTCATAAT-A-TTCGTAAACCAATCAAAAACTCATGTTTTAAATCAATAAAAA
ATACTTAG-', 'seq_2': 'CCCCTGCCGATTTTCGGAGTCAGACCGT-GCGTAATATAAAACGC-CG--GCCC
GCCGATGTATTTGCCGTGGC', 'score': -26, 'identity': 0.3783783783783784, 'matc
h': 28}

seq_1 vs seq_3
----------------------------------------------------------------------
{'seq_1': '-AAACAC-CTCCAGTCATAATATTCGTAAACCAATCAAAAACTCA-TGTTTTAAATCAATAAAAA
AT-ACTTA-G', 'seq_2': 'TCACCACGCCCCCACGACGGAGCTCGT-GGGC-ACCCAGCATTCACTG-CTT-
-ACCACTACGATCTCGCTCACG', 'score': -15, 'identity': 0.4666666666666667, 'matc
h': 35}

seq_1 vs seq_4
----------------------------------------------------------------------
{'seq_1': 'AAACACCTCCAGTCATAATATTCGT-AAACCAATCAAAAACTCAT-GTTTTAAATCAATA-AAAA
ATACTTAG', 'seq_2': 'TTATTTTGAAAGTAACAAT-GGCATCATTATTATTAAATGGT-ATAGGTTT-CTT
GTTTACTGGTATGCTTCA', 'score': -17, 'identity': 0.4246575342465753, 'match':
31}

seq_1 vs seq_5
----------------------------------------------------------------------
{'seq_1': '--AAACACCTCCAGTCATAATATTCGTAAACCAATCA-AAAACTCATGTTTTAAATCAATAAAAA
ATACTTAG', 'seq_2': 'TGATATACTTCAAG-CA-CAT-CACGCAGCCCAACCACTTTTCCCGCGGACGAAG
ACATCTTCCCAGGCAAAC', 'score': -17, 'identity': 0.4246575342465753, 'match':
31}

seq_1 vs seq_6
----------------------------------------------------------------------
{'seq_1': 'AAACACCTCCAGTCATAATA-TTCGTAAACCAAT-CAAAAACTCATGTTTTAAATCAATAAAAAA
TACTTA-G', 'seq_2': '-CTTGCAT-AAGT-ACAACACCGCATAAAATAATACCCGACATAACTAATTCCTT
CATTGGGAGTTTTTTATG', 'score': -13, 'identity': 0.4520547945205479, 'match':
33}

seq_1 vs seq_7
----------------------------------------------------------------------
{'seq_1': '-AAACACCTCCAGTCATAATATTCG--TAAACCAATCAAAAACTCATGTTTTAAATCAATAAAAA
ATACTTAG', 'seq_2': 'GGTTGACAGCGAAT-ACAATACTCGCCGCCAACAGTGTGAAGCTGCCG-CCCGTT
TCTTTGGCGTCAAAGCA-', 'score': -25, 'identity': 0.3698630136986301, 'match':
27}

seq_1 vs seq_8
----------------------------------------------------------------------
{'seq_1': 'AAACACCTCCA-GTCATAATATTC-GTAAACCAATCAAAAACTCATGTTTTAAATCAATAAAAAA
TACTTAG', 'seq_2': 'AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTACA-TCCTGAAATATAT-
TATCATTTAAAGGGGT', 'score': -20, 'identity': 0.3888888888888889, 'match': 2
8}

seq_1 vs seq_9
----------------------------------------------------------------------
{'seq_1': 'AAACACCTCCAGTCATAATATTCG-TAAACCAATCAAAAACTCATGTTTTAAATCAATAAAAAAT
ACTTAG--', 'seq_2': '--TCGCCGAACGTCCAGGCCGTCGCCCTGCCCAGCGCCACCGCCCG-TACCGCCC
```

TGGAACGCCTGCCCAGCC', 'score': -29, 'identity': 0.3424657534246575, 'match':
25}

seq_1 vs seq_10

{'seq_1': 'AAACACCTCCAGTCATAATATTCGTAA-ACCA--ATCAAAAACTCAT-GTTTTA-AATCAATAAA
AAATACTTAG--', 'seq_2': '---CA-CT-GAATCA-ATCCTTCTTAACTTTAGGATCACTTA-TTATGGGG
CTAGGATCAATAAAGGCTATCAAGCA', 'score': -11, 'identity': 0.5194805194805194,
'match': 40}

==========
==========

Alvo: 2
Candidatos: [1, 3, 4, 5, 6, 7, 8, 9, 10]

seq_2 vs seq_1

{'seq_1': 'CCCCTGCCGATTTTCGGAGTCAGACCGT-GCGTAATATAAAACGC-CG--GCCCGCCGATGTATT
TGCCGTGGC', 'seq_2': 'AAAC-ACCTCCAGTCATAAT-A-TTCGTAAACCAATCAAAAACTCATGTTTTAA
ATCAATAAAAAATACTTAG-', 'score': -26, 'identity': 0.3783783783783784, 'matc
h': 28}

seq_2 vs seq_3

{'seq_1': '-C-CC-CTGCCGATTTTCGGAG-TCAGACCGTGCGTAATATAAAACGCCGGCC-CGCCGATGTAT
TTGCCGTGGC', 'seq_2': 'TCACCACGCCCCCACGACGGAGCTC-GTGGGCACCCAGCATTCACTGCTTACC
ACTACGA--T-CTCGCTCACG-', 'score': -19, 'identity': 0.44, 'match': 33}

seq_2 vs seq_4

{'seq_1': 'CCCCTGCCGATTTTCGGAGTCAGACCGTGCGTAATATAAAACG-CCGGCCCGCCGATGTATTTG-
CCG-TGGC', 'seq_2': '-TTATTTTGAAAGTAACAATGGCATCAT-TATTAT-TAAATGGTATAGGTTTCTT
GTTTACTGGTATGCTTCA', 'score': -27, 'identity': 0.3561643835616438, 'match':
26}

seq_2 vs seq_5

{'seq_1': 'CCCCTGCCGAT-TTTCGGAGTCAGA-C-CG-TGCGTAATATAAAACGCCGGCCCGCCGATG-TAT
-TTGCC--G-TGGC', 'seq_2': '----TG-ATATACTTC-AAG-CACATCACGCAGCCCAA-CCACTTTTCC
CG-CGGACGAAGACATCTTCCCAGGCAAAC', 'score': -23, 'identity': 0.468354430379746
83, 'match': 37}

seq_2 vs seq_6

{'seq_1': 'C--CCCT--GCCGATTTTCGGAGTCAGACCGTGCGTAATATAA--AACGCCGGCCCGCCGATGTA
TTTGCCGTGGC', 'seq_2': 'CTTGCATAAGTACAACACCGCA-TAAAATAATACCCGACATAACTAATTCCT
TCATTGGGA-GT-TTT-TTAT-G-', 'score': -24, 'identity': 0.42105263157894735, 'm
atch': 32}

seq_2 vs seq_7

{'seq_1': 'CCCCTG-C--CGATTTTCGGAGTCAGACCGTGCGTAATATAAAACGCCGGCCCGCCGATGTATTT
GCCGT---GGC-', 'seq_2': '-GGTTGACAGCGAATACAATACTC-G-CCGCCAACAGTGT-GAA-G-CTG-
CCGCCCGTTTCTTTGGCGTCAAAGCA', 'score': -15, 'identity': 0.4935064935064935,
'match': 38}

```
seq_2 vs seq_8
--------------------------------------------------------------------
{'seq_1': 'CCCCTGCCGATTTTCGGAGTCAGACCGTGC-GTAATATAAAACGCCGGCCCGCCGATGT-ATTTG
CCGTGGC', 'seq_2': '-AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTACA-TCCTGAAATATAT
TATCATTTAAAGGGGT', 'score': -20, 'identity': 0.3888888888888889, 'match': 2
8}

seq_2 vs seq_9
--------------------------------------------------------------------
{'seq_1': '-CCCCTGCCGATTTTCGGAGTCAGACCGTGCGTAATATAAAACG-CCGGCCCGCCGATGTA--TT
TGCCGTGGC', 'seq_2': 'TCGCCGAACG-TCCAGGCCGTC-G-CCCTGC-CCAGCGCCACCGCCCGTACCGC
CCTGGAACGCCTGCCCAGCC', 'score': -10, 'identity': 0.4864864864864865, 'matc
h': 36}

seq_2 vs seq_10
--------------------------------------------------------------------
{'seq_1': 'CCCCTGCCGAT-TTTCGGAGTCAGACCGT-GCGTAATATAAAACGCCGGC-CCG-CCGATGTA-T
TTGCCGTGGC', 'seq_2': '-CACTG--AATCAATCCTTCTTA-ACTTTAGGATCACTTATTATG-GGGCTAG
GATCAATAAAGGCTATCAAGCA', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

====================================================================================
========================

Alvo: 3
Candidatos: [1, 2, 4, 5, 6, 7, 8, 9, 10]

seq_3 vs seq_1
--------------------------------------------------------------------
{'seq_1': 'TCACCACGCCCCCACGACGGAGCTCGT-GGGC-ACCCAGCATTCACTG-CTT--ACCACTACGAT
CTCGCTCACG', 'seq_2': '-AAACAC-CTCCAGTCATAATATTCGTAAACCAATCAAAAACTCA-TGTTTTA
AATCAATAAAAAAT-ACTTA-G', 'score': -15, 'identity': 0.4666666666666667, 'matc
h': 35}

seq_3 vs seq_2
--------------------------------------------------------------------
{'seq_1': 'TCACCACGCCCCCACGACGGAGCTC-GTGGGCACCCAGCAT-TCACTGCTTACC-ACTACG-ATC
T-CGCTCACG', 'seq_2': '-C-CC-CTGCCGATTTTCGGAG-TCAGACCGTGCGTAATATAAAACGCCGGCC
CGCCGATGTATTTGCCGTGGC-', 'score': -19, 'identity': 0.44, 'match': 33}

seq_3 vs seq_4
--------------------------------------------------------------------
{'seq_1': 'TCACCACG-CCCCCACGACGG-AGC-TCGTGGGCACCCAGCATTCACTGCTT-ACCACTACGATC
TCGCTCACG', 'seq_2': 'TTATTTTGAAAGTAACAATGGCATCATTATTATTAAATGGTATAGGTTTCTTGT
TTACT--GGTAT-GCT-TCA', 'score': -22, 'identity': 0.40540540540540543, 'matc
h': 30}

seq_3 vs seq_5
--------------------------------------------------------------------
{'seq_1': 'TCACCACGCCCCCACGACGGAGCTCG-TGGGCACCCAGCATTCACTGCTTACCACTACGATC-TC
GC--TC-ACG', 'seq_2': 'TGA-TA-TACTTCAAG-CACATCACGCAGCCCAACCA-CTTTTCCCGCGGACG
AAGAC-ATCTTCCCAGGCAAAC', 'score': -11, 'identity': 0.49333333333333335, 'mat
ch': 37}

seq_3 vs seq_6
```

```
----------------------------------------------------------------------
{'seq_1': '-TCAC-CACGCCCCCACGACGGAGCTCGTGGGCACCCAGCATTCACTGCTTAC--CACTACGATC
TCGCTCACG', 'seq_2': 'CTTGCATAAG-TACAACACCGCATAAAAT-AATACCCGACA-TAACTAATTCCT
TCATTGGGA-GTTTTTTATG', 'score': -16, 'identity': 0.44594594594594594, 'matc
h': 33}

seq_3 vs seq_7
----------------------------------------------------------------------
{'seq_1': 'TCACCAC-GCCCCCACGACGGAGCTCGTGGGCACCCAGCATTCA-CTGCTTACCACTACGATCTC
G-CTCA-CG', 'seq_2': 'GGTTGACAGCGAATAC-A-ATA-CTCGCCGCCA-ACAGTGTGAAGCTGCCGCCC
GTTTCTTTGGCGTCAAAGCA', 'score': -14, 'identity': 0.4594594594594595, 'matc
h': 34}

seq_3 vs seq_8
----------------------------------------------------------------------
{'seq_1': '--TCACCACGCCCCCACGACGGAGCTCGTGGGC-ACCCAGCATTCACTG-CTTACCACTACGATC
TCGCTCACG', 'seq_2': 'AGAGACTAATATGTTTTTTCACA-CTTTTCTGCTTTTTTACA-TC-CTGAAATA
-TATTATCATTTAAAGGGGT', 'score': -32, 'identity': 0.33783783783783783, 'matc
h': 25}

seq_3 vs seq_9
----------------------------------------------------------------------
{'seq_1': 'TCACCACGCCCCCACGACGGAGCTCGTGGGCA-C-CCAGCATTCACTGCTTACC-ACTACGATCT
CGCTCA-CG', 'seq_2': 'TCGCCGAACGTCCAGGCCGTCGC-CCTGCCCAGCGCCA-CCGCCCGTACCGCCC
TGGAACG-CCT-GCCCAGCC', 'score': -6, 'identity': 0.5135135135135135, 'match':
38}

seq_3 vs seq_10
----------------------------------------------------------------------
{'seq_1': 'TCAC-CACGCCCCCACGACGGAGCTCGTGG-GCACCCAGCAT-TCACTGCTTACCACTACGATCT
CGCTCACG', 'seq_2': '-CACTGAATCAATC-CTTCTTAACTTTAGGATCACTTATTATGGGGCT-AGGATC
AATAAAGGCTATCAAGCA', 'score': -19, 'identity': 0.410958904109589, 'match': 3
0}

======================================================================
========================

Alvo: 4
Candidatos: [1, 2, 3, 5, 6, 7, 8, 9, 10]

seq_4 vs seq_1
----------------------------------------------------------------------
{'seq_1': 'TTATTTTGAAAGTAACAAT-GGCATCATTATTATTAAATGGT-ATAGGTTT-CTTGTTTACTGGT
ATGCTTCA', 'seq_2': 'AAACACCTCCAGTCATAATATTCGT-AAACCAATCAAAAACTCAT-GTTTTAAAT
CAATA-AAAAATACTTAG', 'score': -17, 'identity': 0.4246575342465753, 'match':
31}

seq_4 vs seq_2
----------------------------------------------------------------------
{'seq_1': '-TTATTTTGAAAGTAACAATGGCATCAT-TATTAT-TAAATGGTATAGGTTTCTTGTTTACTGGT
ATGCTTCA', 'seq_2': 'CCCCTGCCGATTTTCGGAGTCAGACCGTGCGTAATATAAAACG-CCGGCCCGCCG
ATGTATTTG-CCG-TGGC', 'score': -27, 'identity': 0.3561643835616438, 'match':
26}

seq_4 vs seq_3
----------------------------------------------------------------------
```

{'seq_1': 'TTATTTTGAAAGTAACAATGGCATCATTATTATTAAATGGTATAGGTTTCTT--GTTTACTGGTA
T-GCT-TCA', 'seq_2': 'TCACCACG-CCCCCACGACGG-AGC-TCGTGGGCACCCAGCATTCACTGCTTAC
CACTAC-GATCTCGCTCACG', 'score': -22, 'identity': 0.40540540540540543, 'matc
h': 30}

seq_4 vs seq_5

{'seq_1': 'T--TATTTTGAAAGTA-A-CA-ATGGCATCATTATTATTAAATGGTATAGGTTTCTTGTTTACTG
GTATGCTTCA', 'seq_2': 'TGATATACTTCAAGCACATCACGCAGCCCAACCACTTTTCCCGCGGA-CGAAG
ACATCTTCCCAGGCA---AAC-', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

seq_4 vs seq_6

{'seq_1': 'TTATTTTGAAAGTA-ACAATGGCATCATTATTATTAAATGGTATAGGTTTCTTGTTTACTGGTA-
TGCTTCA--', 'seq_2': '--CTTGCATAAGTACAACACCGCAT-A-AAATAATACCCGACATAACTAATTCC
TTCATTGGGAGTTTTTTATG', 'score': -10, 'identity': 0.4864864864864865, 'matc
h': 36}

seq_4 vs seq_7

{'seq_1': '--TT-ATTTTGAAAGTAACAAT-GGCATCATTATTATTAA-ATGGTATAGGTTTCTTGTTTACTG
GTATGCTTCA', 'seq_2': 'GGTTGACAGCGAATACAATACTCGCCGCCAACAGTGTGAAGCTGCCGCCCGTT
TC-T-TTGGC--GT-CAAAGCA', 'score': -17, 'identity': 0.4533333333333333, 'matc
h': 34}

seq_4 vs seq_8

{'seq_1': 'TTATTTTGA-AAGTAACAATGGCA-TCAT-T--ATTATTAAATGGTATAGGTTTCTTGTTTACTG
GTATGCTTCA', 'seq_2': 'AGAGACTAATATGT-TTTTTCACACTTTTCTGCTTTTTTACATCCT-GAAATA
TATT-ATCATTTAAAGGGGT--', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

seq_4 vs seq_9

{'seq_1': 'TTATTTTGAAAGTAACAATGGCATCATTATTATTAAATGGTATAGGTTTCTTGTTTACTGGTATG
-CT--TCA---', 'seq_2': '-T-CGCCGAACGT--CCAGGCCGTC-GCCCTGCCCAGCGCCA-CCGCCCGTA
CCGCCCTGGAACGCCTGCCCAGCC', 'score': -32, 'identity': 0.3684210526315789, 'ma
tch': 28}

seq_4 vs seq_10

{'seq_1': 'TTATTTTGAAAGTAACAATGGCATCATTATTATTAAATGGTAT-AGG-TTTCTTGTTTACTGG-T
ATGCTTCA', 'seq_2': '-CA--CTGAATCAATCCTTCTTAACTTTAGGATCACTTATTATGGGGCTAGGATC
AATAAAGGCTATCAAGCA', 'score': -13, 'identity': 0.4520547945205479, 'match':
33}

==========

Alvo: 5
Candidatos: [1, 2, 3, 4, 6, 7, 8, 9, 10]

seq_5 vs seq_1

{'seq_1': 'TGATATACTTCAAG-CA-CAT-CACGCAGCCCAACCACTTTTCCCGCGGACGAAGACATCTTCCC

AGGCAAAC', 'seq_2': '--AAACACCTCCAGTCATAATATTCGTAAACCAATCA-AAAACTCATGTTTTAAA
TCAATAAAAAATACTTAG', 'score': -17, 'identity': 0.4246575342465753, 'match':
31}

seq_5 vs seq_2

{'seq_1': '----TG-ATATACTTC-AAG-CACATCACGCAGCCCAA-CCACTTTTCCCG-CGGACGAAGACAT
CTTCCCAGGCAAAC', 'seq_2': 'CCCCTGCCGAT-TTTCGGAGTCAGA-C-CG-TGCGTAATATAAAACGCC
GGCCCGCCGATG-TAT-TTGCC--G-TGGC', 'score': -23, 'identity': 0.468354430379746
83, 'match': 37}

seq_5 vs seq_3

{'seq_1': 'TGA-TA-TACTTCAAG-CACATCACGCAGCCCAACCA-CTTTTCCCGCGGACGAAGAC-ATCTTC
CCAGGCAAAC', 'seq_2': 'TCACCACGCCCCCACGACGGAGCTCG-TGGGCACCCAGCATTCACTGCTTACC
ACTACGATC-TCGC--TC-ACG', 'score': -11, 'identity': 0.49333333333333335, 'mat
ch': 37}

seq_5 vs seq_4

{'seq_1': 'TGATATACTTCAAGCACATCACGCAGCCCAACCACTTTTCCCGCGGA-CGAAGACATCTTCCCAG
GCA---AAC-', 'seq_2': 'T--TATTTTGAAAGTA-A-CA-ATGGCATCATTATTATTAAATGGTATAGGTT
TCTTGTTTACTGGTATGCTTCA', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

seq_5 vs seq_6

{'seq_1': 'TGATATACTTCAAGCACATCACGCAGCCCAACCACTTTTCCCG-C-GGACGAA-GACATC-TTCC
CAG---GCAAAC', 'seq_2': '--CT-TGCAT-AAGTACAACAC-C-GCATAA-AATAATACCCGACATAACT
AATTCCTTCATTGGGAGTTTTTTATG', 'score': -17, 'identity': 0.4805194805194805,
'match': 37}

seq_5 vs seq_7

{'seq_1': 'TGATATACTTCAAGCAC-AT-CACGCAGCCCAACCACTTTTCCCGCGGACGAAGACATCTTCCCA
GGCAAA-C-', 'seq_2': '-GGTTGACAGCGAATACAATACTCGCCG-CCAA-CA-GTGTGAAGCTGCCGCCC
GTTTCTTTGGCGTCAAAGCA', 'score': -6, 'identity': 0.5135135135135135, 'match':
38}

seq_5 vs seq_8

{'seq_1': 'TGATA-TACTTCAAG-CACATCACGCAGCCCAACCACTTTTCCCGCGGACGAAGA-CATC-TTCC
CAGGCAAAC', 'seq_2': 'AGAGACTA-AT-ATGTTTTTTCACACTTTTCTGCTTTTTTACATCCTGAAATAT
ATTATCATTTAAAGG--GGT', 'score': -18, 'identity': 0.43243243243243246, 'matc
h': 32}

seq_5 vs seq_9

{'seq_1': 'T-GATATACTTCAAGCACATCACGCAGCCCAACCACTTTTCCCG--CGGACGAAG-ACATCTTCC
CAGGCAAAC', 'seq_2': 'TCGCCGAACGTCCAGGCCGTCGCCCTGCCCAGCGCCACCGCCCGTACCGCCCTG
GAACGCCTGCCCA-GC---C', 'score': -4, 'identity': 0.527027027027027, 'match':
39}

seq_5 vs seq_10

{'seq_1': 'TGATATACTTCAAGCACATC-ACGCAGCCCAACCACTTTTCCCGCGGACGAAGA-C-ATCTTCCC

AGGCAAAC-', 'seq_2': '-CA-CTGAATCAATC-CTTCTTAACTTTAGGATCACTTATTATG-GGGCTAGGA
TCAATAAAGGCTATCAAGCA', 'score': -16, 'identity': 0.44594594594594, 'matc
h': 33}

==========

Alvo: 6
Candidatos: [1, 2, 3, 4, 5, 7, 8, 9, 10]

seq_6 vs seq_1

{'seq_1': '-CTTGCAT-AAGT-ACAACACCGCATAAAATAATACCCGACATAACTAATTCCTTCATTGGGAGT
TTTTTATG', 'seq_2': 'AAACACCTCCAGTCATAATA-TTCGTAAACCAAT-CAAAAACTCATGTTTTAAAT
CAATAAAAAATACTTA-G', 'score': -13, 'identity': 0.4520547945205479, 'match':
33}

seq_6 vs seq_2

{'seq_1': 'CTTGCATAAGTACAACACCGCA-TAAAATAATACCCGACATAACTAATTCCTTCATTGGGA-GT-
TTT-TTAT-G-', 'seq_2': 'C--CCCT--GCCGATTTTCGGAGTCAGACCGTGCGTAATATAA--AACGCCG
GCCCGCCGATGTATTTGCCGTGGC', 'score': -24, 'identity': 0.42105263157894735, 'm
atch': 32}

seq_6 vs seq_3

{'seq_1': 'CTTGCATAAG-TACAACACCGCATAAAAT-AATACCCGACA-TAACTAATTCCTTCATTGGGA-G
TTTTTTATG', 'seq_2': '-TCAC-CACGCCCCCACGACGGAGCTCGTGGGCACCCAGCATTCACTGCTTAC-
-CACTACGATCTCGCTCACG', 'score': -16, 'identity': 0.44594594594594, 'matc
h': 33}

seq_6 vs seq_4

{'seq_1': '--CTTGCATAAGT-ACAA-CACCGCATAAAATAATACCCGACATAACTAATTCCTTCATTGGGAG
TTTTTTATG', 'seq_2': 'TTATTTTGAAAGTAACAATGGCATCAT-TATTATTAAATGGTATAGGTTTCTTG
TTTACTGGTA-TGCTTCA--', 'score': -10, 'identity': 0.4864864864864865, 'matc
h': 36}

seq_6 vs seq_5

{'seq_1': '--CT-TGCAT-AAGTACAACAC-C-GCATAA-AATAATACCCGACATAACTAATTCCTTCATTGG
GAGTTTTTTATG', 'seq_2': 'TGATATACTTCAAGCACATCACGCAGCCCAACCACTTTTCCCG-C-GGACG
AA-GACATC-TTCCCAG---GCAAAC', 'score': -17, 'identity': 0.4805194805194805,
'match': 37}

seq_6 vs seq_7

{'seq_1': '-CTTG-CA-TAAGTACAACAC-CGCATAAAATAATACCCGACATAACTAATTCCTTCATTGGGAG
TTTTTTATG', 'seq_2': 'GGTTGACAGCGAATACAATACTCGCCGCCAACAGT--GTGA-AGCTGCCGCCCG
TTTCTTTGGCG-TCAAAGCA', 'score': -18, 'identity': 0.43243243243243246, 'matc
h': 32}

seq_6 vs seq_8

{'seq_1': 'CTTG-C--ATAAG-TACAACACCGCATAAAATAATACCCGACATAACT-AATTCCTTCATTGGGA
GTTTTTTATG-', 'seq_2': 'AGAGACTAATATGTTTTTTCA-CAC-TTTTCTGCTTTTTTACAT-CCTGAAA

```
T--AT-ATTATCATTTAAAGGGGT', 'score': -26, 'identity': 0.40789473684210525, 'm
atch': 31}

seq_6 vs seq_9
----------------------------------------------------------------------
{'seq_1': 'CTTGCATAA-GTACAACACCGCATAAAATAATACCCGACATAACTAATTCCTTCATTGGGA-G-T
TTTTTA-TG', 'seq_2': '-TCGCCGAACGT-CCAGGCCG-TCGCCCTGCCCAGCGCCACCGCCCGTACC-GC
CCTGGAACGCCTGCCCAGCC', 'score': -24, 'identity': 0.3918918918918919, 'matc
h': 29}

seq_6 vs seq_10
----------------------------------------------------------------------
{'seq_1': 'C-TTGCATAAGTACAACACCGCATAAAATAATACCCGACATAACTAATTCCTTCATTGGGAGTTT
TTTATG--', 'seq_2': 'CACTGAATCAATCCTTCTTAAC-TTTAGGATCACTTATTAT-GGGGCTAGGATCA
AT-AAAGGCTATCAAGCA', 'score': -23, 'identity': 0.3835616438356164, 'match':
28}

======================================================================
========================

Alvo: 7
Candidatos: [1, 2, 3, 4, 5, 6, 8, 9, 10]

seq_7 vs seq_1
----------------------------------------------------------------------
{'seq_1': 'GGTTGACAGCGAAT-ACAATACTCGCCGCCAACAGTGTGAAGCTGCCG-CCCGTTTCTTTGGCGT
CAAAGCA-', 'seq_2': '-AAACACCTCCAGTCATAATATTCG--TAAACCAATCAAAAACTCATGTTTTAAA
TCAATAAAAAATACTTAG', 'score': -25, 'identity': 0.3698630136986301, 'match':
27}

seq_7 vs seq_2
----------------------------------------------------------------------
{'seq_1': '-GGTTGACAGCGAATACAATACTC-G-CCGCCAACAGTGT-GAA-G-CTG-CCGCCCGTTTCTTT
GGCGTCAAAGCA', 'seq_2': 'CCCCTG-C--CGATTTTCGGAGTCAGACCGTGCGTAATATAAAACGCCGGC
CCGCCGATGTATTTGCCGT---GGC-', 'score': -15, 'identity': 0.4935064935064935,
'match': 38}

seq_7 vs seq_3
----------------------------------------------------------------------
{'seq_1': 'GGTTGACAGCGAATAC-A-ATA-CTCGCCGCCA-ACAGTGTGAAGCTGCCGCCCGTTTCTTTGGC
GTCAAAGCA', 'seq_2': 'TCACCAC-GCCCCCACGACGGAGCTCGTGGGCACCCAGCATTCA-CTGCTTACC
ACTACGATCTCG-CTCA-CG', 'score': -14, 'identity': 0.4594594594594595, 'matc
h': 34}

seq_7 vs seq_4
----------------------------------------------------------------------
{'seq_1': 'GGTTGACAGCGAATACAATACTCGCCGCCAACAGTGTGAAGCTGCCGCCCGTTTC-T-TTGGC--
GT-CAAAGCA', 'seq_2': '--TT-ATTTTGAAAGTAACAAT-GGCATCATTATTATTAA-ATGGTATAGGTT
TCTTGTTTACTGGTATGCTTCA', 'score': -17, 'identity': 0.4533333333333333, 'matc
h': 34}

seq_7 vs seq_5
----------------------------------------------------------------------
{'seq_1': '-GGTTGACAGCGAATACAATACTCGCCG-CCAA-CA-GTGTGAAGCTGCCGCCCGTTTCTTTGGC
GTCAAAGCA', 'seq_2': 'TGATATACTTCAAGCAC-AT-CACGCAGCCCAACCACTTTTCCCGCGGACGAAG
ACATCTTCCCAGGCAAA-C-', 'score': -6, 'identity': 0.5135135135135135, 'match':
```

```
38}

seq_7 vs seq_6
----------------------------------------------------------------------
{'seq_1': 'GGTTGACAGCGAATACAATACTCGCCGCCAACAGT--GTGA-AGCTGCCGCCCGTTTCTTTGGCG
-TCAAAGCA', 'seq_2': '-CTTG-CA-TAAGTACAACAC-CGCATAAAATAATACCCGACATAACTAATTCC
TTCATTGGGAGTTTTTTATG', 'score': -18, 'identity': 0.43243243243243246, 'matc
h': 32}

seq_7 vs seq_8
----------------------------------------------------------------------
{'seq_1': 'GGTTGACAGCGAATA-CAATACTCGCCGCCAACAG-TGTGAAGCTGCCGCCCGTTTCTT-TGGCG
TCAA-AGCA', 'seq_2': '---AGAGA-CTAATATGTTTTTTCACACTTTTCTGCTTTTTTACATCCTGAAAT
ATATTATCATTTAAAGGGGT', 'score': -26, 'identity': 0.3783783783783784, 'matc
h': 28}

seq_7 vs seq_9
----------------------------------------------------------------------
{'seq_1': 'GGTTGAC-AGCGAATACAATAC-TCGCCGCCAACAGTGTGAAGCTGCCG-CCCG-TTTCTTTGGC
GT-CAAAGCA', 'seq_2': '--TCGCCGAACG--TCCAGGCCGTCGCCCTGCCCAGCG-CCACCGCCCGTACC
GCCCTGGAACGCCTGCCCAGCC', 'score': -15, 'identity': 0.4666666666666667, 'matc
h': 35}

seq_7 vs seq_10
----------------------------------------------------------------------
{'seq_1': 'GGTTGACAGCGAATACAATACTCGCCGCCAACA-GTGTGAAGCTGCCGCCCGTTTC--T-TTGGC
-GTCAAAGCA', 'seq_2': 'CACTGA-ATC-AAT-C-CTTCTTAACTTTAGGATCACTTATTATGGGGCTAGG
ATCAATAAAGGCTATC-AAGCA', 'score': -15, 'identity': 0.4666666666666667, 'matc
h': 35}

======================================================================
========================

Alvo: 8
Candidatos: [1, 2, 3, 4, 5, 6, 7, 9, 10]

seq_8 vs seq_1
----------------------------------------------------------------------
{'seq_1': 'AGAGACTAATATGTTTTTTCACAC-TTTTCTGCTTTTTTACATCCTGAAATATAT-TATCATTTA
AAGGGGT', 'seq_2': 'AAACACCTCCA-GTCATAATATTCGTAAACCAATCAAAAAC-TCATGTTTTAAATC
AATAAAAAATACTTAG', 'score': -20, 'identity': 0.3888888888888889, 'match': 2
8}

seq_8 vs seq_2
----------------------------------------------------------------------
{'seq_1': '-AGAGACTAATATGTTTTTTCACACTTTTC-TGCTTTTTTACATCCTGAAATATATTATCATTTA
AAGGGGT', 'seq_2': 'CCCCTGCCGATTTTCGGAGTCAGACCGTGCGTAATATAAAAC-GCCGGCCCGCCGA
TGT-ATTTGCCGTGGC', 'score': -20, 'identity': 0.3888888888888889, 'match': 2
8}

seq_8 vs seq_3
----------------------------------------------------------------------
{'seq_1': 'AGAGACTAATATGTTTTTTCACA-CTTTTCTGCTTTTTTACA-TC-CTGAAATA-TATTATCATT
TAAAGGGGT', 'seq_2': '--TCACCACGCCCCCACGACGGAGCTCGTGGGC-ACCCAGCATTCACTG-CTTA
CCACTACGATCTCGCTCACG', 'score': -32, 'identity': 0.33783783783783783, 'matc
h': 25}
```

```
seq_8 vs seq_4
--------------------------------------------------------------------
{'seq_1': 'AGAGACTAATATGT-TTTTTCACACTTTTCTGCTTTTTTACATCCT-GAAATATATT-ATCATTT
AAAGGGGT--', 'seq_2': 'TTATTTTGA-AAGTAACAATGGCA-TCAT-T--ATTATTAAATGGTATAGGTT
TCTTGTTTACTGGTATGCTTCA', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

seq_8 vs seq_5
--------------------------------------------------------------------
{'seq_1': 'AGAGACTA-AT-ATGTTTTTTCACACTTTTCTGCTTTTTTACATCCTGAAATATATTATCATTTA
AAGG--GGT', 'seq_2': 'TGATA-TACTTCAAG-CACATCACGCAGCCCAACCACTTTTCCCGCGGACGAAG
A-CATC-TTCCCAGGCAAAC', 'score': -18, 'identity': 0.43243243243243246, 'matc
h': 32}

seq_8 vs seq_6
--------------------------------------------------------------------
{'seq_1': 'AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTACAT-CCTGAAAT--AT-ATTATCATT
TAAAGGGGT', 'seq_2': 'CTTGCATAA-GTACAACACCGCA-TAAAATAATACCCGACATAACT-AATTCCT
TCATTGGGAGTTTTTTATG-', 'score': -26, 'identity': 0.3783783783783784, 'matc
h': 28}

seq_8 vs seq_7
--------------------------------------------------------------------
{'seq_1': '---AGAGA-CTAATATGTTTTTTCACACTTTTCTGCTTTTTTACATCCTGAAATATATTATCATT
TAAAGGGGT', 'seq_2': 'GGTTGACAGCGAATA-CAATACTCGCCGCCAACAG-TGTGAAGCTGCCGCCCGT
TTCTT-TGGCGTCAA-AGCA', 'score': -26, 'identity': 0.3783783783783784, 'matc
h': 28}

seq_8 vs seq_9
--------------------------------------------------------------------
{'seq_1': 'AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTACATCCTG-AAATATAT-TATCATTTA
AAGGGGT', 'seq_2': '-TCGCCGAACGTCCAGGCCGTCGCCCTGC-CCAGCGCCACCGCCCGTACCGCCCTG
GAACGCCTGCCCAGCC', 'score': -34, 'identity': 0.2916666666666667, 'match': 2
1}

seq_8 vs seq_10
--------------------------------------------------------------------
{'seq_1': '-AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTA-CATCCTG-AAATAT-ATTATCATT
TAAAGGGGT', 'seq_2': 'CACTGAATCA-ATCCTTCTT-A-AC-TTTAGGATCACTTATTATGGGGCTAGGA
TCAATAAAGGCTATCAAGCA', 'score': -14, 'identity': 0.4594594594594595, 'matc
h': 34}

====================================================================================
========================

Alvo: 9
Candidatos: [1, 2, 3, 4, 5, 6, 7, 8, 10]

seq_9 vs seq_1
--------------------------------------------------------------------
{'seq_1': '--TCGCCGAACGTCCAGGCCGTCGCCCTGCCCAGCGCCACCGCCCG-TACCGCCCTGGAACGCCT
GCCCAGCC', 'seq_2': 'AAACACCTCCAGTCATAATATTCG-TAAACCAATCAAAAACTCATGTTTTAAATC
AATAAAAAATACTTAG--', 'score': -29, 'identity': 0.3424657534246575, 'match':
25}
```

```
seq_9 vs seq_2
----------------------------------------------------------------------
{'seq_1': 'TCGCCGAACG-TCCAGGCCGTC-G-CCCTGC-CCAGCGCCACCGCCCGTACCGCCCTGGAACGCC
TGCCCAGCC', 'seq_2': '-CCCCTGCCGATTTTCGGAGTCAGACCGTGCGTAATATAAAACG-CCGGCCCGC
CGATGTA--TTTGCCGTGGC', 'score': -10, 'identity': 0.4864864864864865, 'matc
h': 36}

seq_9 vs seq_3
----------------------------------------------------------------------
{'seq_1': 'TCGCCGAACGTCCAGGCCGTCGC-CCTGCCCA-GC-GCCACCGC-CCGTACCGCCCTGGAACGCC
TGCCCAGCC', 'seq_2': 'TCACCACGCCCCCACGACGGAGCTCGTGGGCACCCAGCATTCACTGCTTACC-A
CTACGATC-TC-GCTCA-CG', 'score': -6, 'identity': 0.5135135135135135, 'match':
38}

seq_9 vs seq_4
----------------------------------------------------------------------
{'seq_1': '-T-CGCCGAACGT--CCAGGCCGTC-GCCCTGCCCAGCGCCA-CCGCCCGTACCGCCCTGGAACG
CCTGCCCAGCC', 'seq_2': 'TTATTTTGAAAGTAACAATGGCATCATTATTATTAAATGGTATAGGTTTCTT
GTTTACTGGTATG-CT--TCA---', 'score': -32, 'identity': 0.3684210526315789, 'ma
tch': 28}

seq_9 vs seq_5
----------------------------------------------------------------------
{'seq_1': 'TCGCCGAACGTCCAGGCCGTCGCCCTGCCCAGCGCCACCGCCCGTACCGCCCTGGAACGCCTGCC
CA-GC---C', 'seq_2': 'T-GATATACTTCAAGCACATCACGCAGCCCAACCACTTTTCCCG--CGGACGAA
G-ACATCTTCCCAGGCAAAC', 'score': -4, 'identity': 0.527027027027027, 'match':
39}

seq_9 vs seq_6
----------------------------------------------------------------------
{'seq_1': '-TCGCCGAACGT-CCAGGCCG-TCGCCCTGCCCAGCGCCACCGCCCGTACC-GCCCTGGAACGCC
TGCCCAGCC', 'seq_2': 'CTTGCATAA-GTACAACACCGCATAAAATAATACCCGACATAACTAATTCCTTC
ATTGGGA-G-TTTTTTA-TG', 'score': -24, 'identity': 0.3918918918918919, 'matc
h': 29}

seq_9 vs seq_7
----------------------------------------------------------------------
{'seq_1': '--TCGCCGAACG--TCCAGGCCGTCGCCCTGCCCAGCG-CCACCGCCCGTACCGCCCTGGAACGC
CTGCCCAGCC', 'seq_2': 'GGTTGAC-AGCGAATACAATAC-TCGCCGCCAACAGTGTGAAGCTGCCG-CCC
G-TTTCTTTGGCGT-CAAAGCA', 'score': -15, 'identity': 0.4666666666666667, 'matc
h': 35}

seq_9 vs seq_8
----------------------------------------------------------------------
{'seq_1': '-TCGCCGAACGTCCAGGCCGTCGCCCTGC-CCAGCGCCACCGCCCGTACCGCCCTGGAACGCCTG
CCCAGCC', 'seq_2': 'AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTACATCCTG-AAATATAT-
TATCATTTAAAGGGGT', 'score': -34, 'identity': 0.2916666666666667, 'match': 2
1}

seq_9 vs seq_10
----------------------------------------------------------------------
{'seq_1': 'TCGCCGAACGTCCAGGCCGTC--GCCCTGCCCAGCGC-CA--CCGCCCGTACCGCCCTGGAACGC
CTGCCCAGCC', 'seq_2': '-CACTGAA--T-CAATCCTTCTTAACTTTAGGATCACTTATTATGGGGCTA-G
GATCAATAAAGGCTATCAAGCA', 'score': -21, 'identity': 0.4266666666666667, 'matc
h': 32}
```

```
================================================================================
========================

Alvo: 10
Candidatos: [1, 2, 3, 4, 5, 6, 7, 8, 9]

seq_10 vs seq_1
--------------------------------------------------------------------
{'seq_1': '---CA-CT-GAATCA-ATCCTTCTTAACTTTAGGATCACTTA-TTATGGGGCTAGGATCAATAAA
GGCTATCAAGCA', 'seq_2': 'AAACACCTCCAGTCATAATATTCGTAA-ACCA--ATCAAAAACTCAT-GTT
TTA-AATCAATAAAAAATACTTAG--', 'score': -11, 'identity': 0.5194805194805194,
'match': 40}

seq_10 vs seq_2
--------------------------------------------------------------------
{'seq_1': '-CACTG--AATCAATCCTTCTTA-ACTTTAGGATCACTTATTATG-GGGCTAGGATCAATAAAGG
CTATCAAGCA', 'seq_2': 'CCCCTGCCGAT-TTTCGGAGTCAGACCGT-GCGTAATATAAAACGCCGGC-CC
G-CCGATGTA-TTTGCCGTGGC', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

seq_10 vs seq_3
--------------------------------------------------------------------
{'seq_1': '-CACTGAATCAATC-CTTCTTAACTTTAGGATCACTTATTATGGGGCT-AGGATCAATAAAGGCT
ATCAAGCA', 'seq_2': 'TCAC-CACGCCCCCACGACGGAGCTCGTGG-GCACCCAGCAT-TCACTGCTTACC
ACTACGATCTCGCTCACG', 'score': -19, 'identity': 0.410958904109589, 'match': 3
0}

seq_10 vs seq_4
--------------------------------------------------------------------
{'seq_1': '-CA--CTGAATCAATCCTTCTTAACTTTAGGATCACTTATTATGGGGCTAGGATCAATAAAGGCT
ATCAAGCA', 'seq_2': 'TTATTTTGAAAGTAACAATGGCATCATTATTATTAAATGGTAT-AGG-TTTCTTG
TTTACTGG-TATGCTTCA', 'score': -13, 'identity': 0.4520547945205479, 'match':
33}

seq_10 vs seq_5
--------------------------------------------------------------------
{'seq_1': '-CA-CTGAATCAATC-CTTCTTAACTTTAGGATCACTTATTATG-GGGCTAGGATCAATAAAGGC
TATCAAGCA', 'seq_2': 'TGATATACTTCAAGCACATC-ACGCAGCCCAACCACTTTTCCCGCGGACGAAGA
-C-ATCTTCCCAGGCAAAC-', 'score': -16, 'identity': 0.44594594594594594, 'matc
h': 33}

seq_10 vs seq_6
--------------------------------------------------------------------
{'seq_1': 'C--AC-TGAAT-CAATCCTTCTTAACTTTAGGATCACTTATTATGGGGCTAGGATCAAT-AAAGG
CTATCAAGCA', 'seq_2': 'CTTGCATAAGTACAACACCGCATAA-AATAATA-CCCGACATA-ACTAATTCC
TTCATTGGGAGTTTTTTATG--', 'score': -23, 'identity': 0.41333333333333333, 'mat
ch': 31}

seq_10 vs seq_7
--------------------------------------------------------------------
{'seq_1': 'CACTGA-ATC-AAT-C-CTTCTTAACTTTAGGATCACTTATTATGGGGCTAGGATCAATAAAGGC
TATC-AAGCA', 'seq_2': 'GGTTGACAGCGAATACAATACTCGCCGCCAACA-GTGTGAAGCTGCCGCCCGT
TTC--T-TTGGC-GTCAAAGCA', 'score': -15, 'identity': 0.4666666666666667, 'matc
h': 35}

seq_10 vs seq_8
```

```
------------------------------------------------------------------
{'seq_1': 'CACTGAATCA-ATCCTTCTT-A-AC-TTTAGGATCACTTATTATGGGGCTAGGATCAATAAAGGC
TATCAAGCA', 'seq_2': '-AGAGACTAATATGTTTTTTCACACTTTTCTGCTTTTTTA-CATCCTG-AAATA
T-ATTATCATTTAAAGGGGT', 'score': -14, 'identity': 0.4594594594594595, 'matc
h': 34}

seq_10 vs seq_9
------------------------------------------------------------------
{'seq_1': '-CACTGAA--T-CAATCCTTCTTAACTTTAGGATCACTTATTATGGGGCTA-GGATCAATAAAGG
CTATCAAGCA', 'seq_2': 'TCGCCGAACGTCCAGGCCGTC--GCCCTGCCCAGCGC-CA--CCGCCCGTACC
GCCCTGGAACGCCTGCCCAGCC', 'score': -21, 'identity': 0.4266666666666667, 'matc
h': 32}

================================================================================
========================
```

c) Utilize o score de cada alinhamento obtido na letra “b” e construir uma matriz de distância entre as 10 OTUs.

Formato da Resposta:

seq_1 seq_2 seq_3 seq_4 seq_5 seq_6 seq_7 seq_8 seq_9 seq_10

seq_1

seq_2

seq_3

seq_4

seq_5

seq_6

seq_7

seq_8

seq_9

seq_10

In [6]:
```
print(distance_matrix)
```

```
[[  0. -26. -15. -17. -17. -13. -25. -20. -29. -11.]
 [-26.   0. -19. -27. -23. -24. -15. -20. -10. -23.]
 [-15. -19.   0. -22. -11. -16. -14. -32.  -6. -19.]
 [-17. -27. -22.   0. -23. -10. -17. -23. -32. -13.]
 [-17. -23. -11. -23.   0. -17.  -6. -18.  -4. -16.]
 [-13. -24. -16. -10. -17.   0. -18. -26. -24. -23.]
 [-25. -15. -14. -17.  -6. -18.   0. -26. -15. -15.]
 [-20. -20. -32. -23. -18. -26. -26.   0. -34. -14.]
 [-29. -10.  -6. -32.  -4. -24. -15. -34.   0. -21.]
 [-11. -23. -19. -13. -16. -23. -15. -14. -21.   0.]]
```

d) Implemente o algoritmo UPGMA (unweighted pair group method with arithmetic mean) para realizar a construção da árvore filogenética das 10 OTUs consideradas na letra “a”

Entrada do algoritmo: matriz de distâncias construída no item “c”

In [3]:
```
print(distance_matrix)
np.savetxt(f'{dir}/dist_matrix.txt', distance_matrix, fmt='%d')
```

```
[[  0. -26. -15. -17. -17. -13. -25. -20. -29. -11.]
 [-26.   0. -19. -27. -23. -24. -15. -20. -10. -23.]
 [-15. -19.   0. -22. -11. -16. -14. -32.  -6. -19.]
 [-17. -27. -22.   0. -23. -10. -17. -23. -32. -13.]
 [-17. -23. -11. -23.   0. -17.  -6. -18.  -4. -16.]
 [-13. -24. -16. -10. -17.   0. -18. -26. -24. -23.]
 [-25. -15. -14. -17.  -6. -18.   0. -26. -15. -15.]
 [-20. -20. -32. -23. -18. -26. -26.   0. -34. -14.]
 [-29. -10.  -6. -32.  -4. -24. -15. -34.   0. -21.]
 [-11. -23. -19. -13. -16. -23. -15. -14. -21.   0.]]
```

Saída: a árvore filogenética

Os índices das sequências começam em 0.

seq_1 -> 0,

seq_2 -> 1,

...

In [10]:
```
import upgma
upgma.run(distance_matrix)
```

Out[10]:
```
(4, ((6, ((3, (8, 7)), (1, 0))), (2, (9, 5))))
```