5. Phylogenetic trees

- The tree of life

- The tree, an abstract object

- **Building an array of distances**

- The UPGMA algorithm

- Differences are not always what they look like

- The diversity of bioinformatics algorithms

- The application domains in microbiology

Arrays or matrices?

- A matrix is a mathematical object
  - Linear algebra

- An array is a data structure
  - A matrix may be implemented as an array
  - But an array is not always a matrix

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

	1	2	3
1			
2			
3			

1) …AACGTATAGCATTAGCTAGT…
2) …AACCTATTGCCATTAGTTAGT…
3) …AACGTATAGCATTAGCTACT…

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

	1	2	3
1			
2			
3			

1) …AACGTATAGCATTAGCTAGT…
2) …AACCTATTGCCATTAGTTAGT…
3) …AACGTATAGCATTAGCTACT…

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

	1	2	3
1	**0**		
2			
3			

1) ...AACGTATAGCATTAGCTAGT...
2) ...AACCTATTGCCATTAGTTAGT...
3) ...AACGTATAGCATTAGCTACT...

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

	1	2	3
1	0	15	
2			
3			

1) ...AACGTATAGCATTAGCTAGT...
2) ...AACCTATTGCCATTAGTTAGT...
3) ...AACGTATAGCATTAGCTACT...

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

	1	2	3
1	0	15	8
2			
3			

1) …AACGTATAGCATTAGCTAGT…
2) …AACCTATTGCCATTAGTTAGT…
3) …AACGTATAGCATTAGCTACT…

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

1) ...AACGTATAGCATTAGCTAGT...
2) ...AACCTATTGCCATTAGTTAGT...
3) ...AACGTATAGCATTAGCTACT...

	1	2	3
1	0	15	8
2	15	0	
3			

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

	1	2	3
1	0	15	8
2	15		
3			

1) …AACGTATAGCATTAGCTAGT…
2) …AACCTATTGCCATTAGTTAGT…
3) …AACGTATAGCATTAGCTACT…

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file
  2. Compute the distances between this sequence and all the other sequences of the file
  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

1) ...AACGTATAGCATTAGCTAGT...
2) ...AACCTATTGCCATTAGTTAGT...
3) ...AACGTATAGCATTAGCTACT...

	1	2	3
1	0	15	8
2	15	0	
3			

Filling an array of distances

- The input: several genomic sequences in a file

  1. Read a sequence in the file

  2. Compute the distances between this sequence and all the other sequences of the file

  3. Store the values of the distances in the array

  Repeat the process with the next sequence in the file

1) ...AACGTATAGCATTAGCTAGT...
2) ...AACCTATTGCCATTAGTTAGT...
3) ...AACGTATAGCATTAGCTACT...

	1	2	3
1	0	15	8
2	15	0	24
3			

Filling an array of distances

- The input: several genomic sequences in a file

    1. Read a sequence in the file

    2. Compute the distances between this sequence and all the other sequences of the file

    3. Store the values of the distances in the array

    Repeat the process with the next sequence in the file

	1	2	3
1	0	15	8
2	15	0	24
3	8	24	0

1) ...AACGTATAGCATTAGCTAGT...
2) ...AACCTATTGCCATTAGTTAGT...
3) ...AACGTATAGCATTAGCTACT...