Sequence Alignment with Dot Matrix & Visualization

Dot Matrix Method

Compare two sequences by plotting one along the **x-axis** and the other along the **y-axis**. Place a **dot** wherever characters match. Continuous diagonal lines represent regions of similarity.

Think of "connect-the-dots." A perfect diagonal = two sequences are nearly identical. Breaks = mismatches or gaps.

Interpreting Dot Plots

1. **Perfect diagonal line**: high similarity (identical or closely related sequences).
2. **Parallel diagonals**: repeats or duplications.
3. **Broken diagonal**: mutations (insertions, deletions, substitutions).

Example:

- Sequence 1: **ATGCT**
- Sequence 2: **ATGAT**

Plot shows 3 dots in a row (ATG match), then break (mutation).

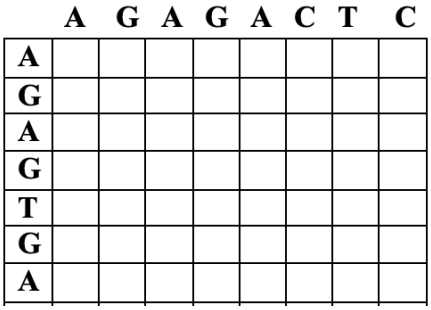
Related concepts for Dot Plot:

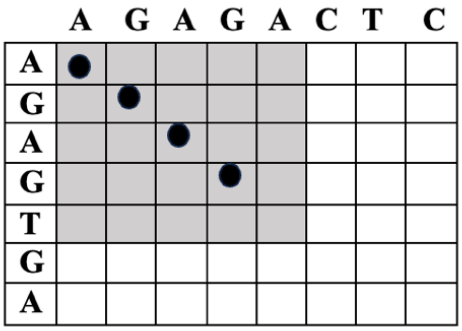
1. Window size – Number of nucleotides compare each time. (usually odd number)
2. Stringency – the minimum number of nucleotides in the window must be 'match' so that a dot can be placed.
3. Mismatch limit – the maximum number of nucleotides in the window can be 'not match', so that a dot can still be placed.
4. Mismatch limit = Window size – Stringency

Example 1: Compare the following sequences and find the region of similarity between two sequences. (window size = 5, stringency = 3)

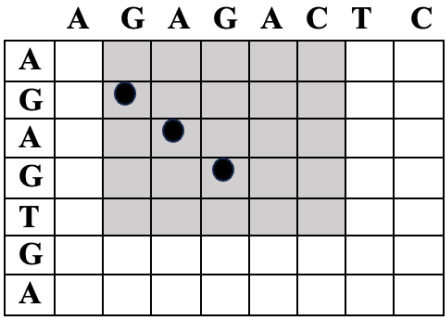
AGAGACTC
AGAGTGTG

Step 1:

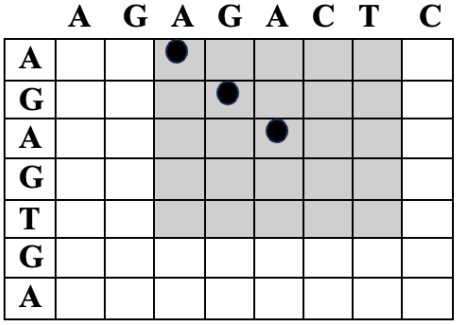




Step 2 and 3:

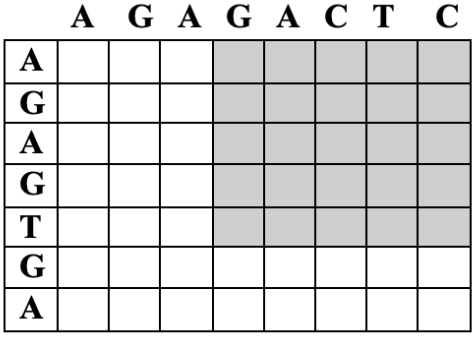


3 matches



3 matches

Step 4:



Only 2 matches

Example 2: Compare the following sequences and find the region of similarity between two sequences. (window size = 5, stringency = 3)

TGACCATGG
GGTACCAGC

Alignment Visualization Tools

- Software/tools:

    1. **Dotlet (Java-based)** – interactive dot plot.
    2. **NCBI BLAST Viewer** – graphical alignment.
    3. **UCSC Genome Browser** – large-scale sequence visualization.

- Importance: Makes raw alignment data understandable.

Practical Session: Dot Plot Practice

Generate and interpret dot plots.

1. Use **Dotlet** (or online dot plot tool).
2. Input two short DNA sequences, e.g.:

    Seq1: ATGCTTAGC

    Seq2: ATGCGTAGC

3. Observe dot plot → diagonal with small mismatches.
4. Modify one sequence (add insertion) → see how dot breaks.

task:

- Generate dot plot for human vs mouse insulin gene.
- Identify conserved (similar) regions.

📝 Quick Review Questions

1. What does a continuous diagonal line in a dot plot represent?
2. How can you detect insertions or deletions using dot plots?
3. Why are visualization tools important in bioinformatics?