Lecture – Sequence analysis

Global Alignment and Local Alignment

Dot Matrix Method

Sequence analysis

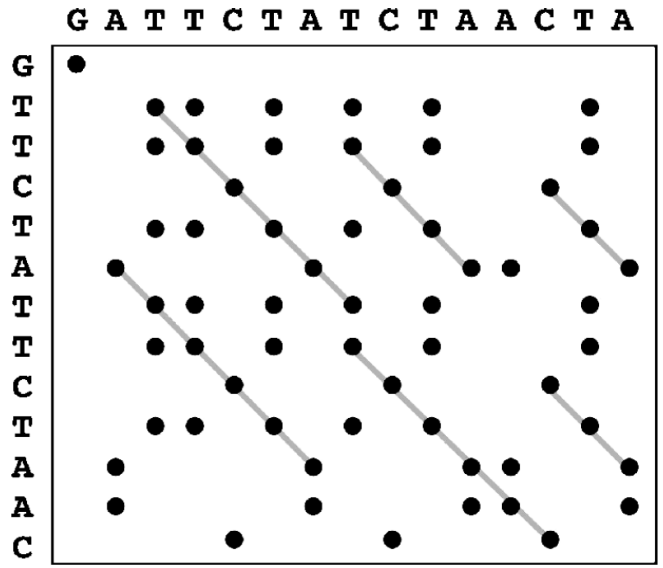
Global Alignment and Local Alignment

Alignment algorithms

Sequence alignment Dot Matrix Method

- The most basic sequence alignment method is the dot matrix method, also known as the dot plot method.
- It is a graphical way of comparing two sequences in a two-dimensional matrix.
- In a dot matrix, two sequences to be compared are written in the horizontal and vertical axes of the matrix.
- The comparison is done by scanning each residue of one sequence for similarity with all residues in the other sequence.
- If a residue match is found, a dot is placed within the graph.
- Otherwise, the matrix positions are left blank.
- When the two sequences have substantial regions of similarity, many dots line up to form contiguous diagonal lines, which reveal the sequence alignment.
- If there are interruptions in the middle of a diagonal line, they indicate insertions or deletions.

Dot Matrix



- Parallel diagonal lines within the matrix represent repetitive regions of the sequences.

Source: Essential Bioinformatics by Jin Xiong

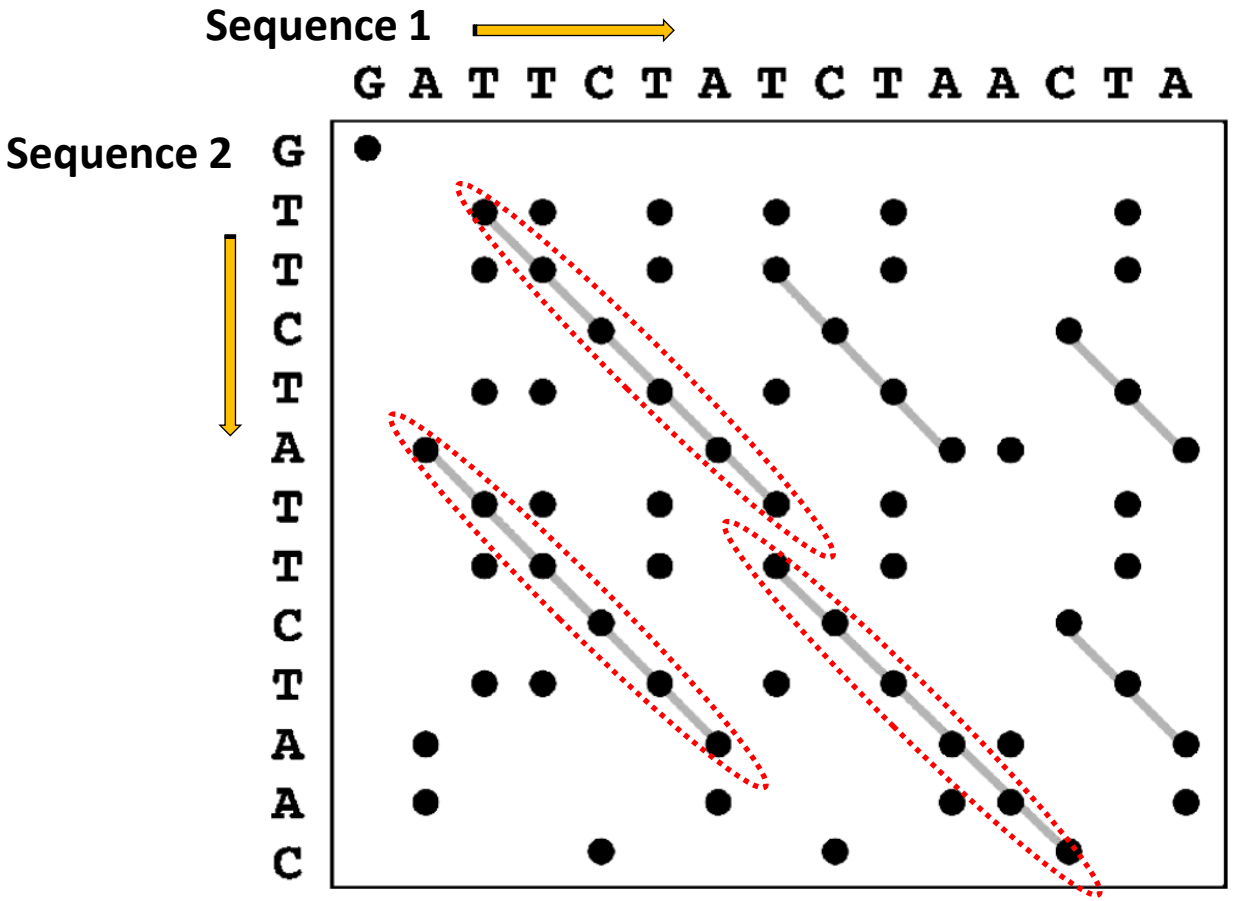
Sequence analysis

Global Alignment and Local Alignment

Alignment algorithms

Sequence alignment

Dot Matrix Method



- ✓ Lines linking the dots in diagonals indicate sequence alignment.
- ✓ Diagonal lines above or below the main diagonal represent internal repeats of either sequence.

Source: Essential Bioinformatics by Jin Xiong

Sequence alignment

Dot Matrix Method

Limitation?

- A problem exists when comparing large sequences using the dot matrix method, namely, the high noise level.
- In most dot plots, dots are plotted all over the graph, obscuring identification of the true alignment.
- For DNA sequences, the problem is particularly acute because there are only four possible characters in DNA and each residue therefore has a one-in-four chance of matching a residue in another sequence.

Source: Essential Bioinformatics by Jin Xiong

Sequence alignment

Dot Matrix Method ⬇

Limitation?

- The dot matrix method displays all possible sequence matches.
- However, it is often up to the user to construct a full alignment with insertions and deletions by linking nearby diagonals.
- Another limitation of this visual analysis method is that it lacks statistical rigor in assessing the quality of the alignment.
- The method is also restricted to pairwise alignment.
- It is difficult for the method to scale up to multiple alignment.

Source: Essential Bioinformatics by Jin Xiong

Sequence alignment — Dot Matrix Method

How it was made better?

A "window" of fixed length used as:

- To reduce noise, instead of using a single residue to scan for similarity, a filtering technique has to be applied, which uses a "window" of fixed length covering a stretch of residue pairs.
- When applying filtering, windows slide across the two sequences to compare all possible stretches.
- Dots are only placed when a stretch of residues equal to the window size from one sequence matches completely with a stretch of another sequence.
- This method has been shown to be effective in reducing the noise level.
- The window is also called a tuple.

Window size or tuple

- The window size can be manipulated so that a clear pattern of sequence match can be plotted.
- However, if the selected window size is too long, sensitivity of the alignment is lost.

Source: Essential Bioinformatics by Jin Xiong

Sequence alignment Dot Matrix Method ⬇

Variations and applications?

- ✓ There are many variations of using the dot plot method.
- ✓ For example, a sequence can be aligned with itself to identify internal repeat elements.
- ✓ In the self comparison, there is a main diagonal for perfect matching of each residue.
- ✓ If repeats are present, short parallel lines are observed above and below the main diagonal.

Note: For comparing protein sequences, a weighting scheme has to be used to account for similarities of physicochemical properties of amino acid residues.

Source: Essential Bioinformatics by Jin Xiong

Sequence alignment Dot Matrix Method

Variations and applications?

- ✓ The dot matrix method gives a direct visual statement of the relationship between two sequences and helps easy identification of the regions of greatest similarities.
- ✓ One particular advantage of this method is in identification of sequence repeat regions based on the presence of parallel diagonals of the same size vertically or horizontally in the matrix.
- ✓ The method thus has some applications in genomics.
- ✓ It is useful in identifying chromosomal repeats and in comparing gene order conservation between two closely related genomes.
- ✓ It can also be used in identifying nucleic acid secondary structures through detecting self-complementarity of a sequence (for example, those that form the stems of a hairpin structure – can also be identified using a dot plot).

Source: Essential Bioinformatics by Jin Xiong

Sequence alignment Dot Matrix Method

Webservers implementing dot matrix:

- ✓ **The following are examples of webservers that provide pairwise sequence comparison using dot plots.**
  - ✓ **Dotmatcher and Dottup: two programs of the EMBOSS package**
  - ✓ **Dothelix: has option for length threshold (similar to window size)**
  - ✓ **MatrixPlot: the program uses colored grids to indicate alignment**