

Math Minute 2.1 What Can You Learn from a Dot Plot?

- Suppose that two sequences are identical except that a segment is inverted in one sequence, relative to the other. Explain how such an inversion would be displayed in a dot plot.

The inversion would create a black streak in the dot plot, running diagonally from top right to bottom left. An example is shown below, obtained by comparing the protein query in Discovery Question 10 in Chapter 1 to the same sequence with amino acids 8–20 reversed. This dot plot was made with dotplot.xls.



Other examples can be seen in Figures 2.4 and 2.20. Note that inversions are not identifiable in the sliding window view because the sequences are no longer similar along a diagonal segment from top left to bottom right, the order in which the sliding window measures similarity.

- Suppose that two sequences are identical except that a segment is deleted from the middle of one sequence and not from the other. Explain how such a deletion would be displayed in a dot plot.

In this situation, there would be a diagonal segment of black, then a horizontal jump (if deletion in sequence labeling the columns) or a vertical jump (if deletion in sequence labeling the rows), after which the diagonal segment would pick up again. An example is shown below, obtained by comparing the sequence from Discovery Question 10 in Chapter 1 to the same sequence with amino acids 8–20 deleted in the sequence labeling the rows. Unlike inversions (see Math Minute Discovery Question 1) deletions are identifiable in both the identity and sliding window similarity dot plots.

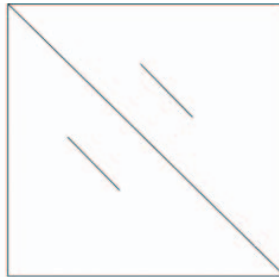


3. What would be the value of using a dot plot to compare a sequence to a second sequence, as well as the reverse complement of that second sequence?

By comparing to both the original sequence and the reverse complement of the sequence, homology could be detected on both strands. For example, if one sequence codes for a gene, and the other is the complementary strand for a homologous gene, this homology could still be detected.

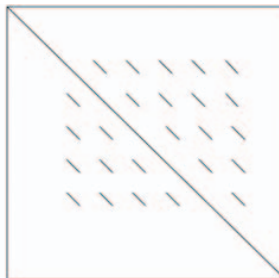
4. Dot plots can detect many interesting sequence features by using the exact same sequence on both the horizontal and vertical axes. Sketch the dot plot of a 100 kb sequence in which a 20 kb segment is duplicated.

The dot plot would be symmetric, with a solid diagonal line (because the sequence is identical to itself). The repeat shows up as lines, above and below the main diagonal, and 20% the length of the main diagonal.



5. Sketch the dot plot of a 1 kb sequence in which a motif of approximately 50 consecutive bases appears six times in the sequence.

The properties of this sketch are similar to the sketch in Math Minute Discovery Question 4, but with shorter sequences repeated more often. Multiple repeats result in a checkerboard type pattern such as the following:



6. What would be the value in using a dot plot to compare a sequence to its own reverse complement?

Basic RNA structure can be detected, since hairpin loops show up as identical on a dot plot comparing a sequence to its own reverse complement. The bottom two plots in dotplot.xls allow students to explore this possibility. For example, the following plot of identity was made by comparing the sequence `ACGTGGCCATATATCGCCACGT` to itself. Notice that the last 7 nt are complementary to the first 7 nt, which can be seen by the diagonal strip in the sliding window plot.

