Answers

Jichang Zhang

2022/7/1

4. For record similarities and differences between two amplicon sequences, there are a lot of metrics can be used to measure similarity.

The easiest way is to count the differences between two sequences (Hamming distance). We can also divide it by the whole length of the amplicon sequences.

Besides, we can calculate Euclidean distance between two sequences, it measures the distance of the two vectors in a 2-D space. The smaller the distance, the more similar the two sequences are.

What’s more, we can also calculate the correlation angle of the two sequences, it measures the angle between the two vectors. The smaller the value, the more similar the two sequences are.

5. To compare two amplicon sequences, we can do a pairwise comparison, it would involve the mutation matrix, which take evolutionary process into consideration. Different mutation (transversion vs. transition) types cause different penalties.

What’s more, if the two amplicon sequences are multiple measurements, we can use the consensus profile instead of the consensus sequence to do multiple sequences alignment. This would keep the information of mutation in each amplicon and show more details.