在遗传生物学的研究中，序列的同源性是一个非常重要的研究目标，对其的研究在揭示生命本质、探索生物进化等多个方面都有着极其重要的意义。我们设计了一个模型对RNA序列的遗传距离进行测算，以此达到对序列同源性进行量化评估的目的。在对该模型进行准确度和复杂度的分析之后，我们设计了一个有效算法来确定多个RNA序列的祖先序列，并建立出相应的进化树。

在模型一中，我们将RNA的碱基序列转化为字符串，以编辑距离算法为原理，采取循环、矩阵和动态规划相结合的方法，建立了算法1.1来求解两条序列的Levenstein距离，基于序列长度的不同可能会对结果产生影响的考虑，我们使用公式 对Levenstein距离进行归一化处理，得出两条序列的GLD距离。通过与MEGA软件数据进行比较，得出了的最合适值为1.64的结论。最后基于对算法1.1的复杂性评估结果，我们优化了模型，建立了算法1.2

在模型二中，我们使用软件MEGA的数据作为标准，对模型一进行了准确性评估。我们将突变率c定义为突变数占原始序列总长度的百分比。首先固定突变率，改变序列长度，得出突变率一定时模型精确率会随序列长度变化而波动，但准确率不低于94%，平均值为96.84%。然后固定序列长度，改变突变率，得出随着突变率变化精确性呈现先减小后增大的趋势，且总体精确率不低于60%。最后使用新冠病毒的实例进行检验。经过以上分析，我们的结论为模型一适用于近种生物的同源性分析与遗传距离测算，且准确性较高。

在模型三中，我们使用非加权组平均法对样本序列组进行了进化树的构建。然后以最长公共子序列算法为基础建立模型，将求得的样本序列组的最长公共子序列与所有样本序列对齐后，以最大概率的原则对突变点位进行数据填补，以此确定多个样本序列的祖先序列

综上所述，我们建立了序列距离测算、进化树构建和祖先序列构建三个模型。经过分析和实例测试，结果表明，我们的模型具有较高的精确度。同时由于算法的空间复杂性是线性的，我们的模型在内存固定的情况下可以处理更长的序列。

In the study of genetic biology, sequence homology is a very important research goal, and the study of sequence homology is of great significance in revealing the nature of life and exploring biological evolution. We designed a model to measure the genetic distance of RNA sequences, so as to achieve the purpose of quantitative evaluation of sequence homology. After analyzing the accuracy and complexity of the model, we designed an efficient algorithm to determine the ancestry of multiple RNA sequences and build corresponding evolutionary trees.

In Model 1, we convert the RNA base sequence into a string. Based on the principle of editing distance algorithm, algorithm 1.1 is established to solve the Levenstein distance of two sequences by combining the method of cycle, matrix and dynamic programming. Considering that different sequence lengths may affect the results, We use formula to normalized processing of Levenstein distance, and conclude the GLD distance of the two sequances. By comparison with MEGA software data, the most suitable value of is 1.64. Finally, based on the complexity evaluation results of algorithm 1.1, we optimized the model and established algorithm 1.2

In Model 2, we use the data of software MEGA as the standard to evaluate the accuracy of model 1. We define mutation rate as the percentage of mutations in the total length of the original sequence. First, we assume that the mutation rate is fixed and the sequence length was changed. It is concluded that when the mutation rate is fixed, the model accuracy rate will fluctuate with the sequence length, but the accuracy rate was no less than 94%, with an average of 96.84%. Then we assume that the length of the sequence is fixed and the mutation rate is changed. The accuracy shows a trend of decreasing first and then increasing with the change of mutation rate, and the overall accuracy was not less than 60%. Finally, an example of the COVID-19 is used for test the model. After the above analysis, we conclude that model 1 is suitable for homology analysis and genetic distance calculation of the near species, and has high accuracy.

In Model 3, we use the longest common subsequence algorithm to build the model, and fill the data of mutation points with the principle of maximum probability, so as to determine the ancestor sequence of multiple sample sequences. Then the unweighted group average method is used to construct the evolutionary tree of the sample sequences. (to be modified)

To sum up, we establish sequence distance measurement, evolutionary tree construction and ancestral sequence construction models. Through analysis and case test, the results show that our model has high accuracy. At the same time, because the spatial complexity of the algorithm is linear, our model can handle longer sequences with fixed memory.