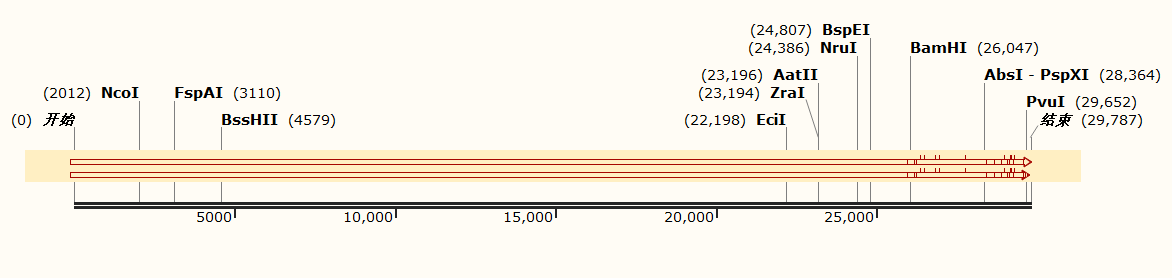
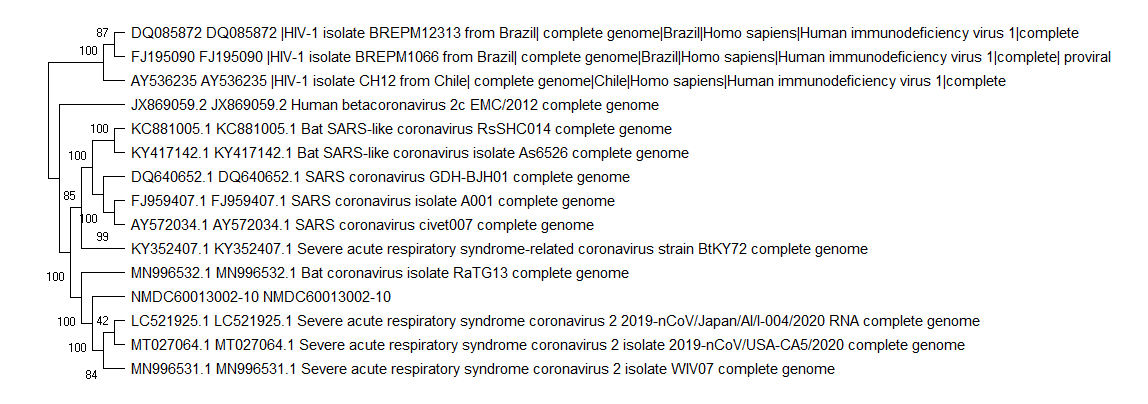
冠状病毒前期工作：

通过对新冠病毒序列与SARS、SHC014蝙蝠冠状病毒序列进行比对，可以确定新冠病毒不是由此二种病毒人工构建嵌合体形成的。在新冠病毒序列当中没有发现新的酶切位点，不太可能是人工重组形成。



接下来通过构建系统发生树，试图通过不同区段序列建树结果中发生的拓扑学结构变化推断重组的可能，但目前由于建树进展相对缓慢，且建树流程不一致导致导致进展缓慢，之后我们会和专门负责建树的同学沟通，获得相对统一的序列集合以及建树方法，将此步骤进一步完善。

前期构建的全基因组序列系统发生树：



之后按任务要求将新冠病毒 序列通过BLAST方法同HIV、流感病毒以及猪瘟病毒进行序列对比，发现匹配结果基本均为短序列，且大多处于蛋白质的保守结构域内。以下以HIV病毒比对结果为例：

以下为新型病毒与HIV病毒进行BLAST搜索到的匹配

>lcl|Query\_5510:44-67 KC156122 |HIV-1 isolate CH256\_TF from Malawi| complete genome|Malawi|Homo sapiens|Human immunodeficiency virus 1|complete

ATTTGTGGGTTTATAACACACAAG NPS13超家族区段

>lcl|Query\_5510:5135-5174 KC156122 |HIV-1 isolate CH256\_TF from Malawi| complete genome|Malawi|Homo sapiens|Human immunodeficiency virus 1|complete

GTTTTACAGACATCACTATGAAAGCAGACATCCAAAAATA orf1ab内

>lcl|Query\_5510:9102-9125 KC156122 |HIV-1 isolate CH256\_TF from Malawi| complete genome|Malawi|Homo sapiens|Human immunodeficiency virus 1|complete

ATTTGTGGGTTTATAACACACAAG NPS13超家族区段

>lcl|Query\_5513:6822-6837 MH705150 |HIV-1 isolate 20-02 from Cameroon| complete genome|Cameroon|Homo sapiens|Human immunodeficiency virus 1|complete

ACAATGTATACATTAA orf1ab内

>lcl|Query\_5513:3688-3707 MH705150 |HIV-1 isolate 20-02 from Cameroon| complete genome|Cameroon|Homo sapiens|Human immunodeficiency virus 1|complete

ACAAATGATATAAGACAATT ORF1内

>lcl|Query\_5513:1425-1444 MH705150 |HIV-1 isolate 20-02 from Cameroon| complete genome|Cameroon|Homo sapiens|Human immunodeficiency virus 1|complete

CTTTACAAGTACTAAAGGAA

>lcl|Query\_5509:4450-4472 AY703910 |HIV-1 isolate 04ZASK143B1 from South Africa| complete genome|South Africa|Homo sapiens|Human immunodeficiency virus 1|complete

AAAGCAAAAATCATTAAAGACTA 在ORF1内

>lcl|Query\_5509:1820-1834 AY703910 |HIV-1 isolate 04ZASK143B1 from South Africa| complete genome|South Africa|Homo sapiens|Human immunodeficiency virus 1|complete

GCTTTATCAAAGTAA 在ORF1内

>lcl|Query\_5509:6065-6079 AY703910 |HIV-1 isolate 04ZASK143B1 from South Africa| complete genome|South Africa|Homo sapiens|Human immunodeficiency virus 1|complete

ATTGTACAAATGTTA NSP11序列

>lcl|Query\_5512:5665-5679 DQ093597 |HIV-1 isolate 03ZASK223B1 from South Africa| complete genome|South Africa|Homo sapiens|Human immunodeficiency virus 1|complete

AGGAATTGTCAACAA

S蛋白

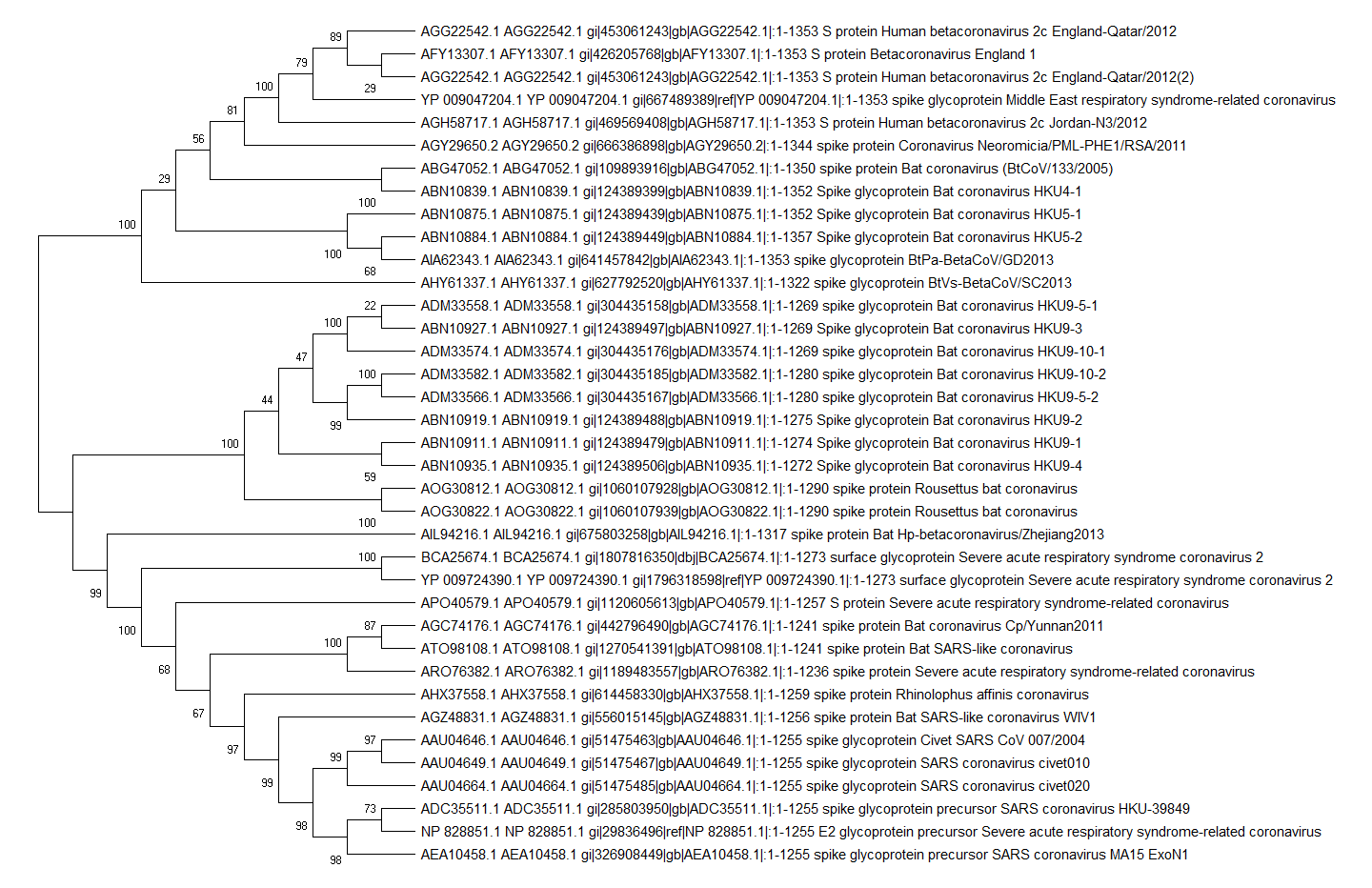
故不认为新冠病毒的序列来源于任务要求比对的病毒，之后将研究范围初步定在冠状病毒科内。

对各种冠状病毒进行了全基因组的simplot分析，发现病毒同源性在不同区段之间差异显著。

基因组层面上与新冠病毒最接近的RATG13在S蛋白序列上与新冠病毒差异反而较大。但由于全基因组水平的图像不够直观，所以接下来可能进行分区段具体的画图。

考虑到不止一篇文献强调S蛋白结构对于病毒宿主范围的影响是显著的，因此进行的S蛋白氨基酸序列的系统发育树构建。但S蛋白形成的系统树和病毒宿主范围并不完全吻合。

S蛋白氨基酸序列系统发生树：



The evolutionary history was inferred by using the Maximum Likelihood method based on the Poisson correction model [1]. The bootstrap consensus tree inferred from 1000 replicates [3] is taken to represent the evolutionary history of the taxa analyzed [3]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [3]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 37 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1001 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2].

单倍型分析：目前单倍型分析能够做出其他文献的结果，但没有独创性，之后单倍型相关的工作可能会围绕高频突变区域进行，在根据高频突变区域进行系统发育树的构建。

或者对编码区讨论氨基酸序列的影响。下一步将对S蛋白的核苷酸序列进行进化树分析并重点与其他病毒进行比较。并且根据侵染所用受体对进化树进行标注。

下一步准备从密码子偏好性角度对感染不同宿主的冠状病毒进行分析。希望能够找出不同宿主冠状病毒之间密码子的偏好性差异，与新冠病毒比较，推测其可能来源或中间宿主。

此外，阅读文献得知冠状病毒重组比较常见，因此应该从进化率和重组率角度进行分析。

将重组时间分析与系统分析结合。

有文献认为新冠病毒S蛋白RBD是通过复杂演化而来而非突变的积累，之后可以寻找相关证据。

接下来一周内将学习使用滑动窗口分析和错配、中性检测分析。