WORLD ORGANISATION FOR ANIMAL HEALTH
Protecting animals, preserving our future

OIE Regional Virtual Training Workshop on Molecular Epidemiology Techniques (MET) for Rabies

19th-20th October 2021

Manuals

Changchun Veterinary Research Institute, Changchun China

http://cvri.caas.cn/

Contents

I. Amplification of N and G genes 1

Introduction 1

Materials 1

Procedure 2

Results 4

II. Sequence identity calculation using BioEdit® 6

Introduction 6

Installation 7

Application: sequence comparison 7

III. Phylogenetic analysis using MEGA® 12

Introduction 12

Download 12

Installation 13

Phylogenetic analysis using NJ method 13

I. Amplification of N and G genes

Introduction

The N and G genes of rabies virus (RABV) are 1,353 bp and 1,575 bp in length (**Fig 1.1**), and encode structural nucleoprotein (N protein) and glycoprotein (G protein). N protein is an essential component of the nucleocapsid that wraps the RNA genome; G protein forms individual spikes on the surface of the viral envelope, with functions in both virus attachment and penetration. Usually, the N gene is used for phylogenetic analysis to inspect the relationship of different virus strains. This protocol shows how to amplify the complete N and G genes from positive brain tissue samples for sequence analysis.

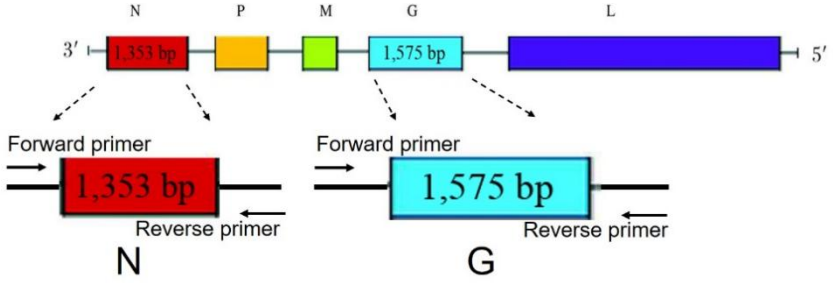


Fig 1.1 Genome structure of RABV showing the location of N and G genes as well as amplicon sizes.

Materials

Primer pairs of N and G genes

Name	Sequence	Location
NF (forward primer)	CCGGATCCAAGAAAAAACAG	12-31
NR (reverse primer)	ATCGATCGAAAGGAGGGGTG	1478-151
GF (forward primer)	CGCTGCATTTYATCAAAGTCAA	3221-324
GR (reverse primer)	TGMAYGGAGTTCAAGGAGGAC	4979-499

Reagents:
RNA extraction kit, SuperScript® III First Strand Synthesis System (Invitrogen, USA, catalog no. 18080051), Phusion® High-Fidelity DNA Polymerase (New England BioLabs, USA, catalog no. M0530), dNTPs, ddH_2O

Note: All PCR reagents are stored at -20°C prior to use and placed on ice during the procedure.

Other materials and equipment:
0.2ml MicroAmp® reaction tubes, pipettes (10-1000µL), vortex mixer, microcentrifuge, PCR system, microwave oven, gel mould, conical flask (250ml), agarose gel, molecular weight ladder, gel electrophoresis tank / power pack, gel documentation system (such as GelDoc-It®), measuring cylinder, UV light device

Procedure

1. Extract total RNA using a commercial kit.

2. Prepare the reaction mix in 0.2ml MicroAmp® reaction tubes as follows, keeping all reagents on ice.

3. Reverse transcription procedure

Reagent	Volume per reaction (µL)
Random Hexamers (50ng/µL)	1µL
10mM dNTP mix	1µL
RNA	8µL

65°C for 5min → then place on ice for at least 1min.

Reagent	Volume per reaction (µL)
10×RT Buffer	2µL
25mM $MgCl_2$	4µL
0.1M DTT	2µL
RNase OUT® (40 U/µL)	1µL
SuperScript® III RT (200 U/µL)	1µL

25°C for 10min → 50°C for 50min → 85°C for 5min.

Add 1µL of RNase H → 37°C for 20min.

4. PCR procedure

Note: N and G gene amplification use the same PCR mix but with different primers.

Reagent	**Volume Per Reaction (μL)**
Phusion DNA polymerase	1
5×Buffer	8
Forward primer (20pmol/L)	0.6
Reverse primer (20pmol/L)	0.6
dNTPs	2
ddH_2O	26.8
cDNA	1
Total	**40**

5. Mix tube contents gently using a pipette and then close the lids tightly; label the lids clearly. Transfer tubes to the PCR machine.

6. PCR programs

N gene:

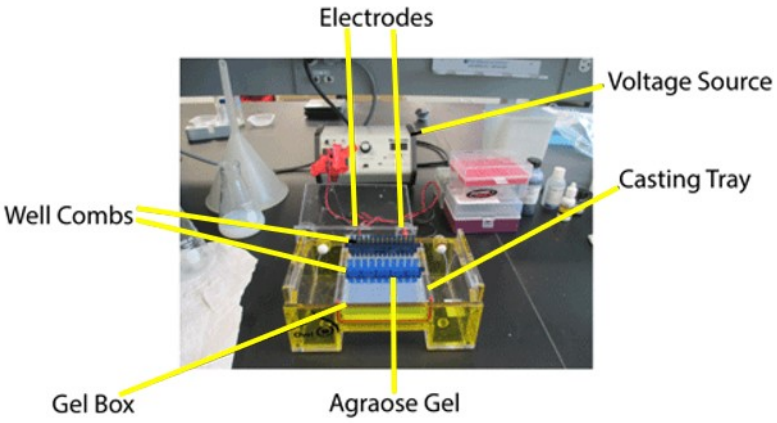
94°C for 2 min	1 Cycle
94°C for 30s	30 Cycles
48°C for 40s	
72°C for 90s	
72°C for 10 min	1 Cycle
Hold at 4°C	

G gene:

94°C for 2 min	1 Cycle
94°C for 30s	40 Cycles
56.7°C for 30s	
72°C for 40s	
72°C for 10 min	1 Cycle
Hold at 4°C	

7. Gel electrophoresis

To enable loading of a DNA sample, it must be mixed with a suitable volume of gel loading buffer. A DNA size marker must also be run alongside the samples.



i. Prepare and pour a 1% agarose gel with EtBr or alternative solution in a gel tray with well comb in place; once the gel has set, remove well comb.

ii. Carefully load a molecular weight ladder into the first lane of the gel.

iii. Carefully load the samples into the other wells of the gel.

iv. Run the gel at 80-150V until the dye line is approximately 75-80% of the way down the gel.

v. Turn off power, disconnect the electrodes from the power source, and then carefully remove the gel from the gel box.

vi. Using a UV light device, visualize your DNA fragments.

Results

The expected results from electrophoresis are shown in **Fig 1.2**. Amplicons of N and G genes should be at 1,498 bp and 1,778 bp, respectively. Their positions can be roughly read against the Marker. Negative samples will not contain such amplicons. These amplicons can be used directly in Sanger sequencing.

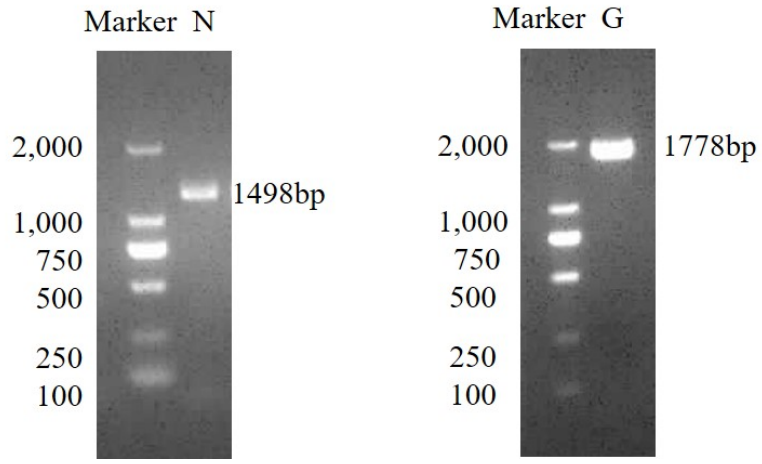


Fig 1.2 Electrophoresis analyses of PCR amplification of N and G genes.

II. Sequence identity calculation using BioEdit®

Introduction

BioEdit® v7.2 can be freely downloaded from software.informer (https://bioedit.software.informer.com). **Fig 2.1** shows the page from which to download the software by clicking the green 'Download' button. BioEdit® includes the tools required by scientists and technicians to perform specific sequence analyses – such as sequence alignment, ABI sequencer data tracing or RNA analysis.

Here, the function of sequence comparison—used to calculate identity or similarity of two homologic sequences—will be introduced.

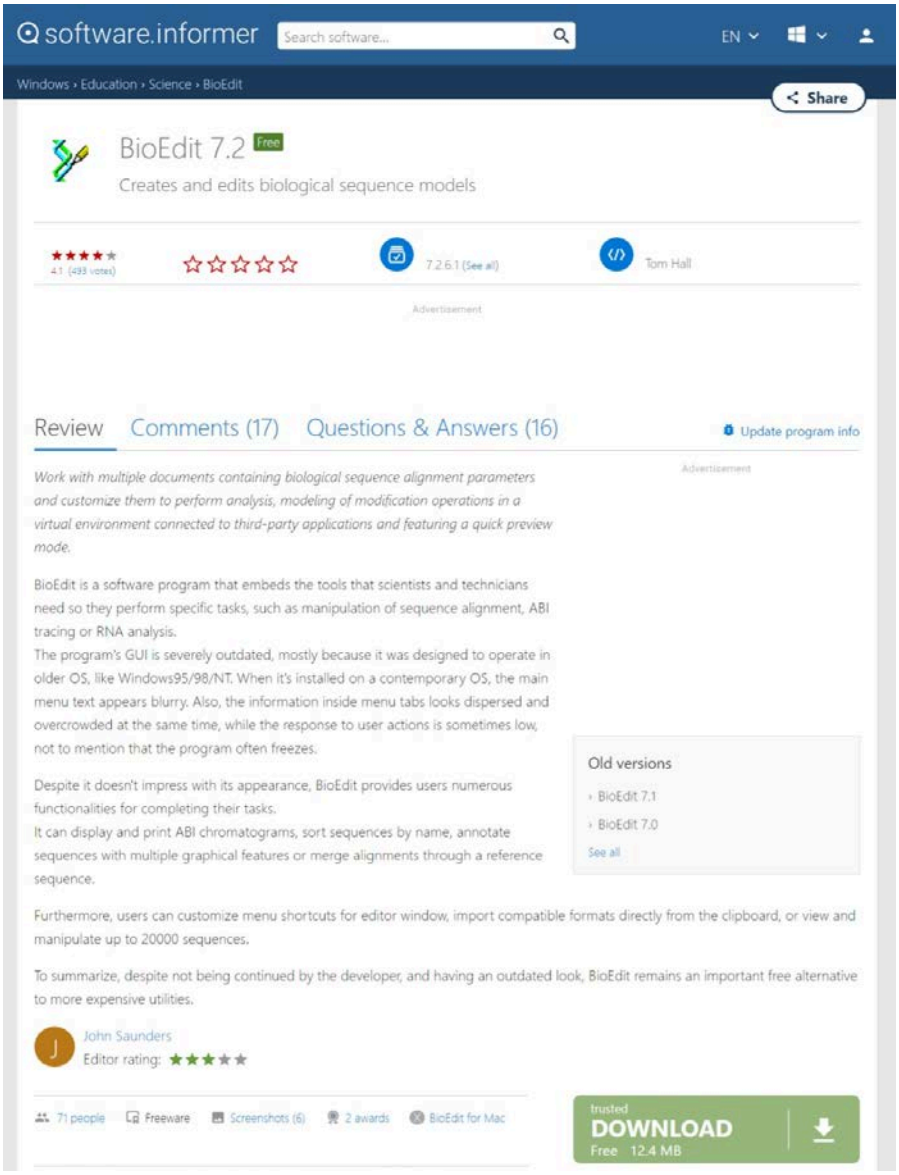


Fig 2.1 BioEdit® software download page.

Installation

Once the bioedit.zip file is downloaded, unzip the package and double click the 'setup.exe' icon. Proceed through the installation window prompts. The software will be self-installed with default settings that are optimal for almost all applications.

Application: sequence comparison

1. Open the software from the START ("BioEdit" application file in the downloaded folder), which will open a new BioEdit Sequence Alignment Editor (**Fig. 2.2**).

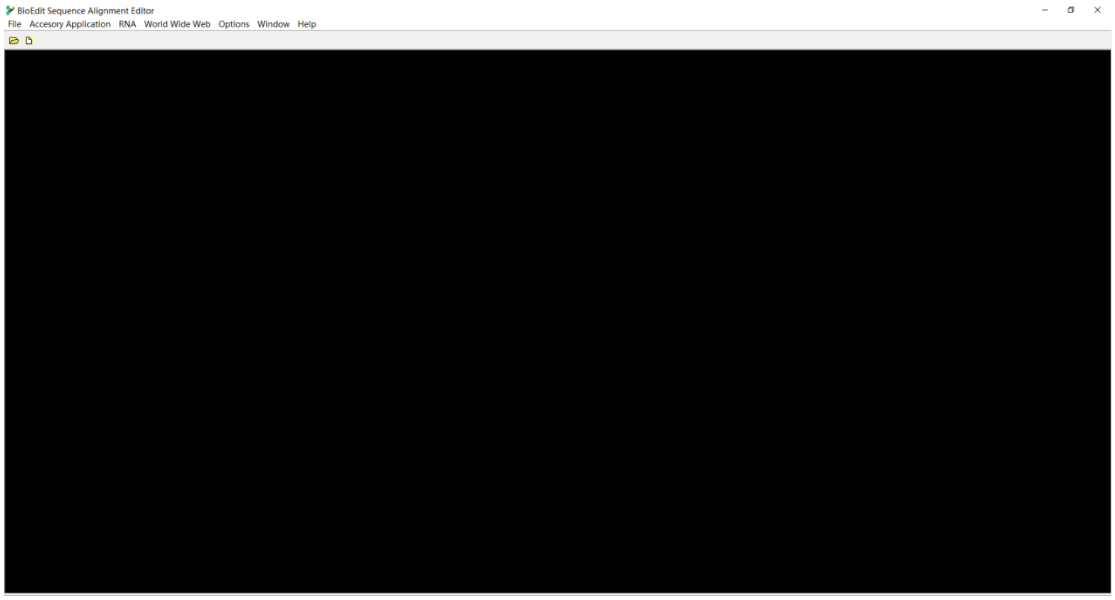


Fig. 2.2 BioEdit Sequence Alignment Editor window.

2. Open sequence file

   a) Open sequence file from 'File → Open', and a dialog box will pop up (**Fig. 2.3**). BioEdit can read a lot of sequence file formats, such as Genbank (*.gbk, *.gen, *.gb, *.gnk) and Fasta (*.fas, *.fst, *.fsa, *.fasta) files.

   b) Here I will demonstrate with the Fasta format using rabies virus (RABV) glycoprotein encoding sequences that were downloaded previously from GenBank. Select the Fasta target file and open it. The BioEdit® editor will contain two columns: a sequence title column (on the left) and sequence column (on the right) (**Fig. 2.4**).

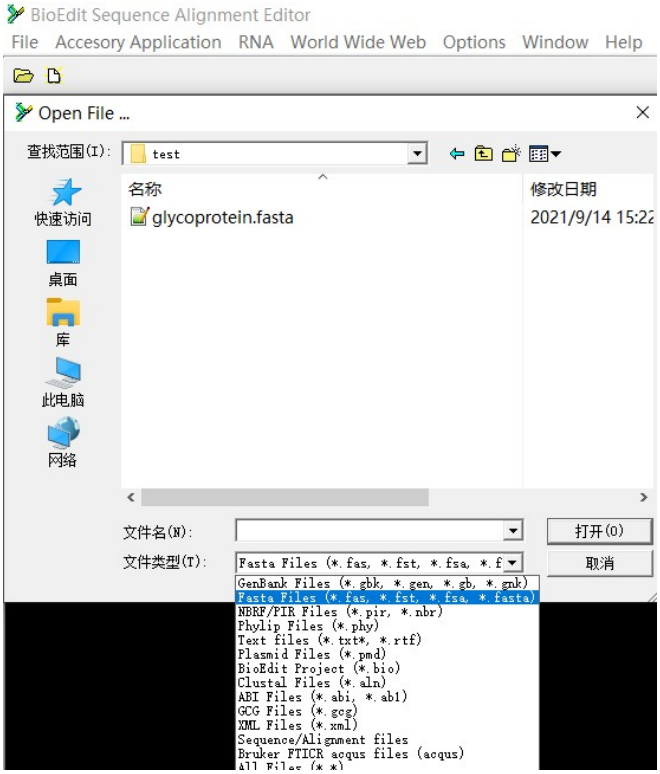


Fig. 2.3 Dialog box for opening a sequence file.

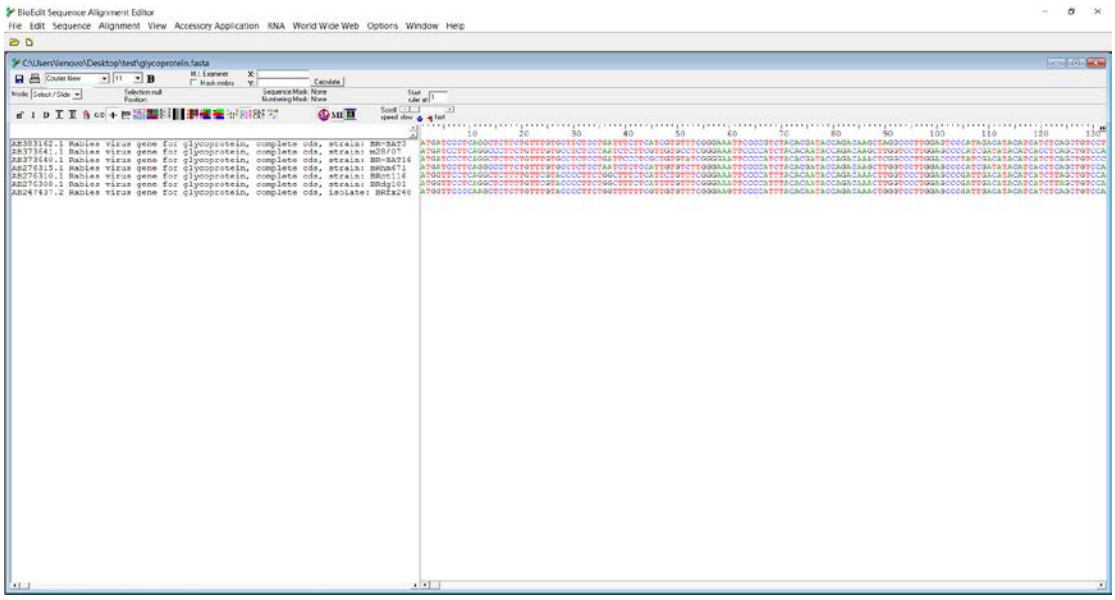


Fig. 2.4 Example BioEdit® editor window containing seven rabies viral glycoprotein sequences.

3. Multiple sequence alignment

   a) Sequentially click 'Edit' → 'Select All Sequences' or just press Ctrl+A to select all sequences.

   b) Then go to 'Accessory Application' → 'ClustalW Multiple alignment', and a dialog box of ClustalW Options will pop up (**Fig. 2.5**). Here we tick 'Full

Multiple alignment' and 'Bootstrap NJ Tree', and specify 1000 for the number of bootstraps. These options are also default settings.

c) Then, click the 'Run ClustalW' button to activate the ClustalW module, and a logger window will pop up to show the process of clustalw alignment (**Fig. 2.6**).

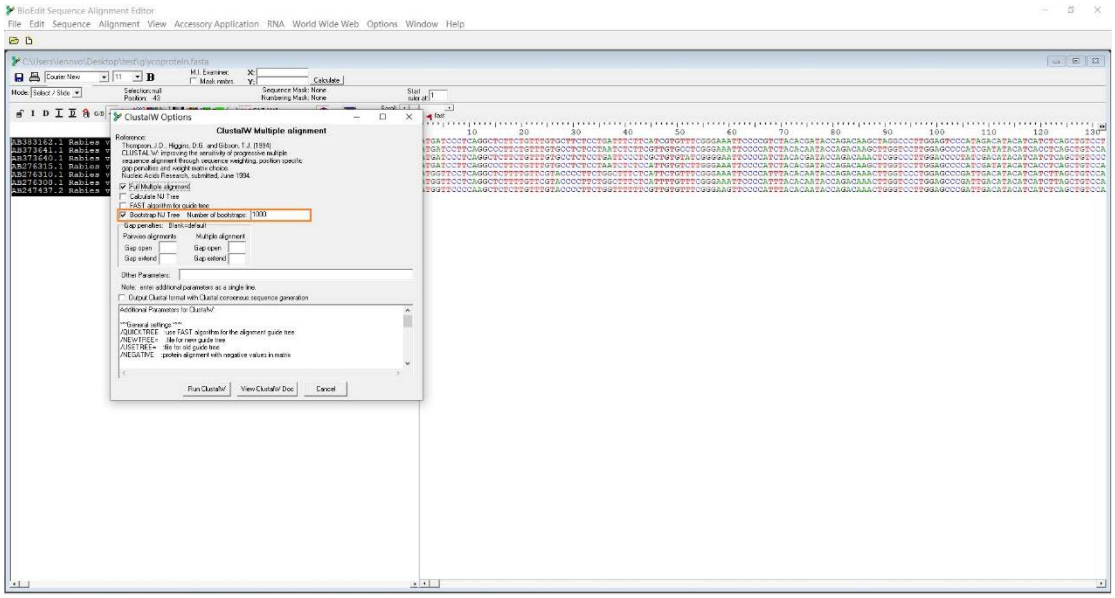


Fig. 2.5 The clustalw options for multiple alignment.

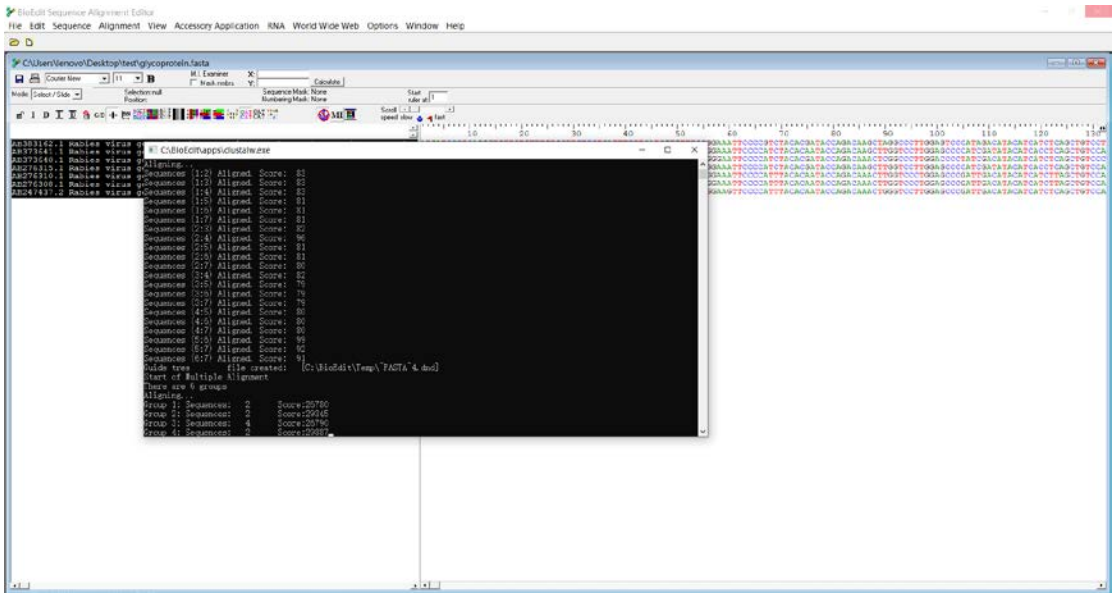


Fig. 2.6 A logger window pops up to show the process of clustalw alignment.

d) The clustalW alignment will take a while, depending on the size of data to be analyzed and the computer performance. When completed, a new untitled editor window will appear that shows the aligned sequences (**Fig. 2.7**). Keep the untitled editor on the top layer so that any further operations will apply to

the aligned sequences.

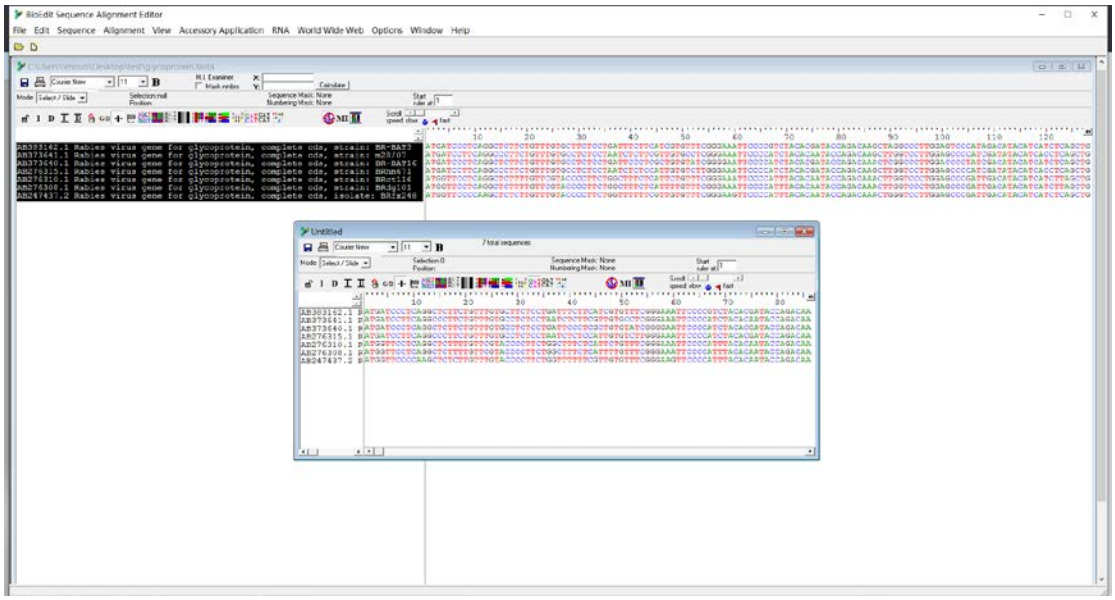


Fig. 2.7 The aligned sequences will appear in a new untitled editor window.

4. Save identity matrix

   a) Go to 'Alignment → Sequence Identity Matrix' and save the identity matrix as a tab delimited file (.txt) or comma delimited table (.csv).

   b) Here we save it as a csv table (note that no comma is included in the sequence title). This table can be opened using Excel (**Fig. 2.8**).

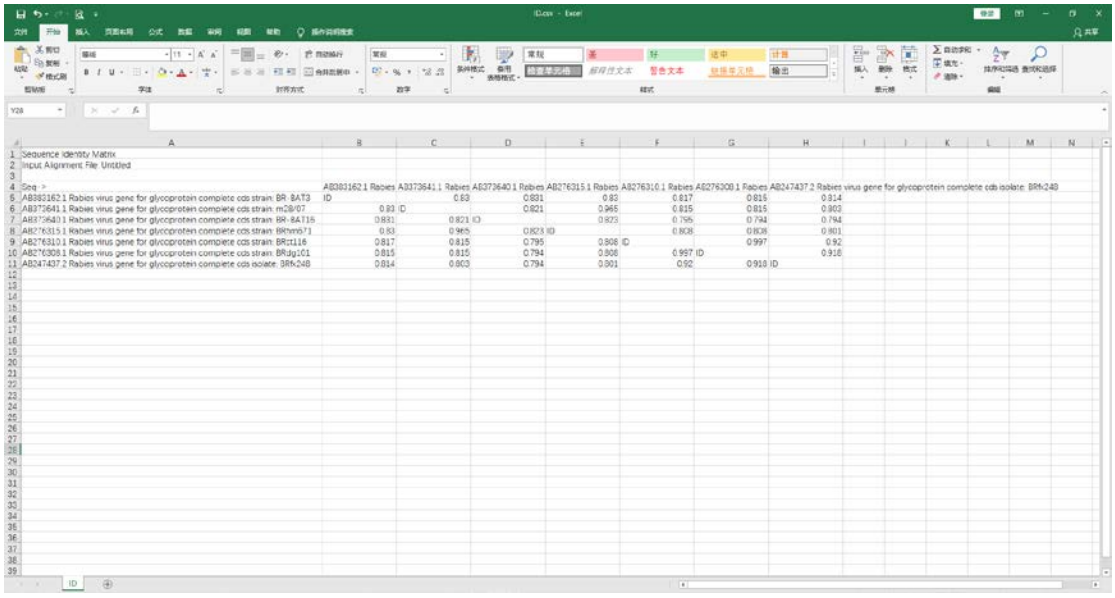


Fig. 2.8 The calculated identity matrix of sequences.

5. Review output

   a) From this matrix we can know the identity between any sequence pairs. In

the example shown, the identity between strains BR-BAT3 and m28/07 is 0.83; this means BR-BAT3 is 83% nucleotide identical or similar to m28/07.

III. Phylogenetic analysis using MEGA®

Introduction

Currently, many computer programs are available for estimating evolutionary distances and reconstructing phylogenetic trees from molecular data. However, most of them are written for specific methods and cannot be interconnected easily because of their inflexible input and output file formats. MEGA® software represents an interactive, user-friendly platform for estimating evolutionary distances, reconstructing phylogenetic trees, and computing basic statistical quantities that are of evolutionary interest.

Here, a neighbor-joining (NJ) phylogenetic analysis of seven rabies viral glycoprotein sequences will be demonstrated.

Download

MEGA® can be freely accessed at https://www.megasoftware.net (**Fig. 3.1**). The end-user should download (via the green “Download” button) the correct version according to the computer OS (selected from the dropdown boxes).

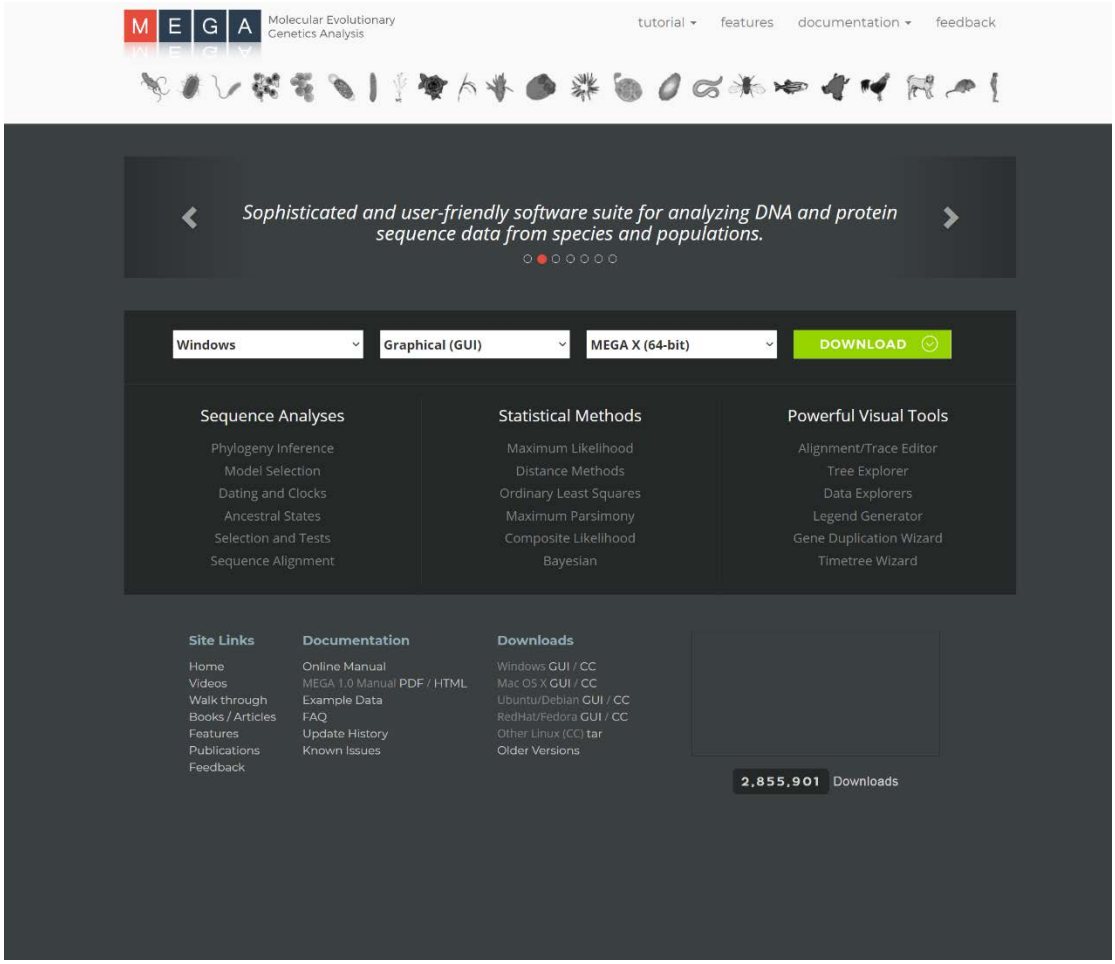


Fig. 3.1 Homepage of MEGA® software.

Installation

Proceed through the installation window prompts. Once downloaded, just double click the package. The software will be self-installed with default settings that are optimal for almost all applications.

Phylogenetic analysis using NJ method

1. Open the software
   a) Double click the MEGA icon (or application file in the downloaded folder) to start the software (**Fig. 3.2**)

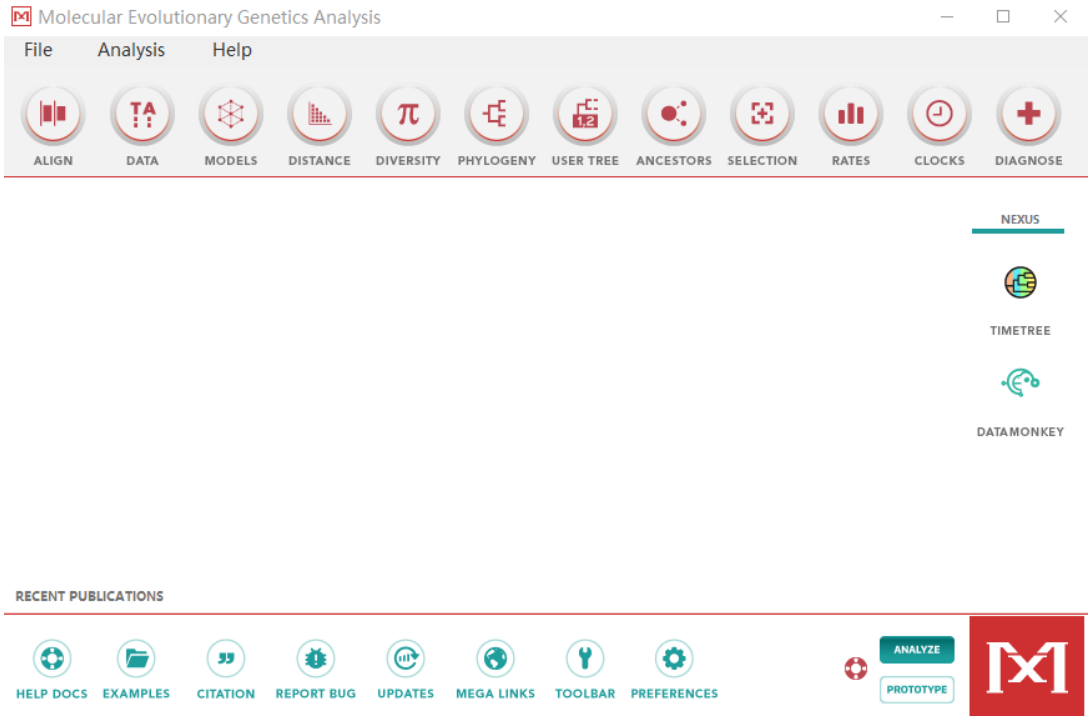


Fig. 3.2 The main page of MEGA.

2. Open data file

   a) Go to 'File' → 'Open a File/Session' to open a Fasta data file.

   b) When you select the target file, a dialog box will appear and ask 'How would you like to open this fasta file?' (**Fig. 3.3**). Click 'Align', and an alignment explorer will open the sequence target (**Fig. 3.4**).

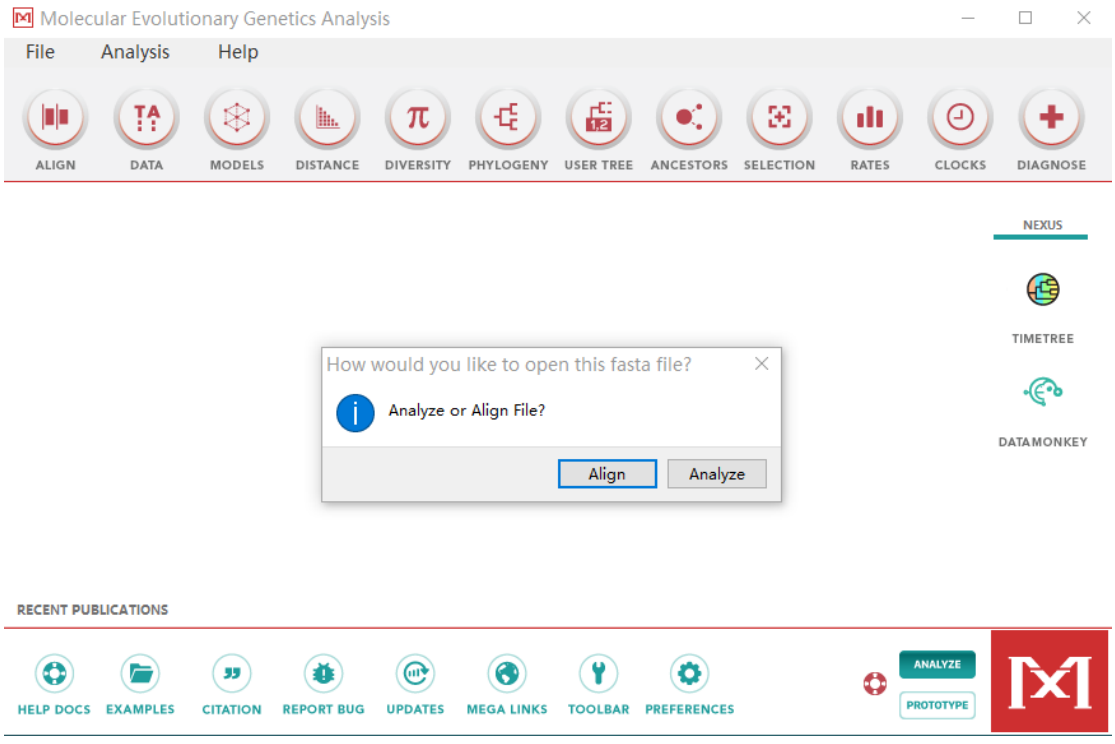


Fig. 3.3 Dialog box to select 'Align' or 'Analyze'.

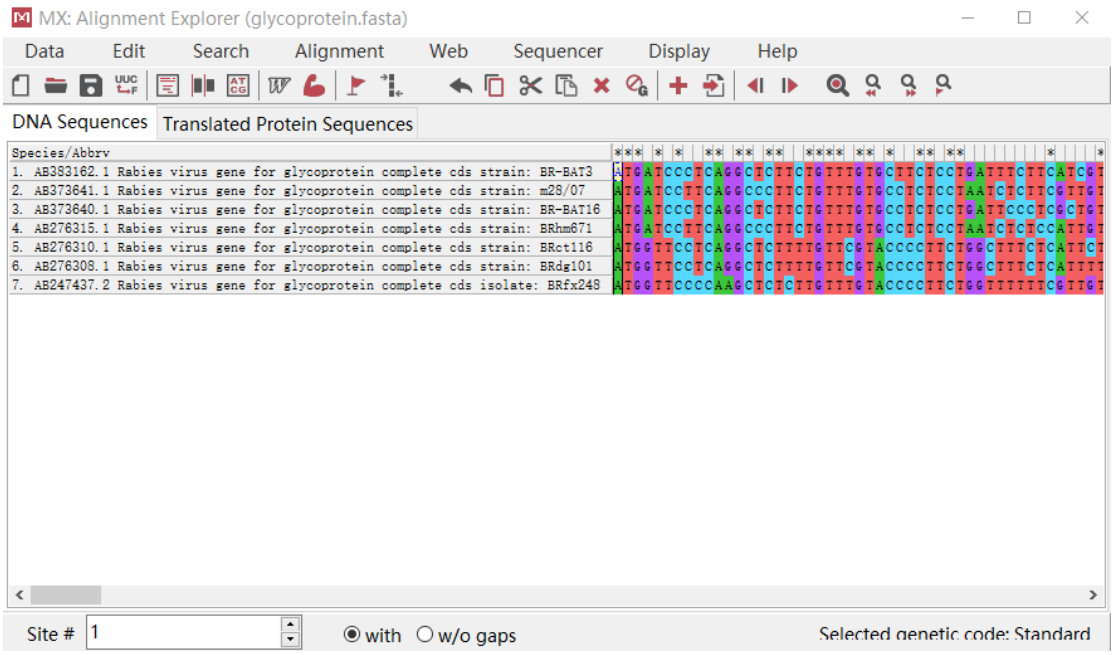


Fig. 3.4 A fasta file opened using MEGA® alignment explorer.

3. Multiple sequence alignment

   a) Go to 'Edit' → 'Select All' or press 'Ctrl+A' to select all sequences.

   b) Then go to 'Alignment' → 'Align by ClustalW' or just click the 'w' icon below the menu to activate the alignment function.

   c) A box will appear to ask for ClustalW options (**Fig. 3.5**). Here we use the default settings.

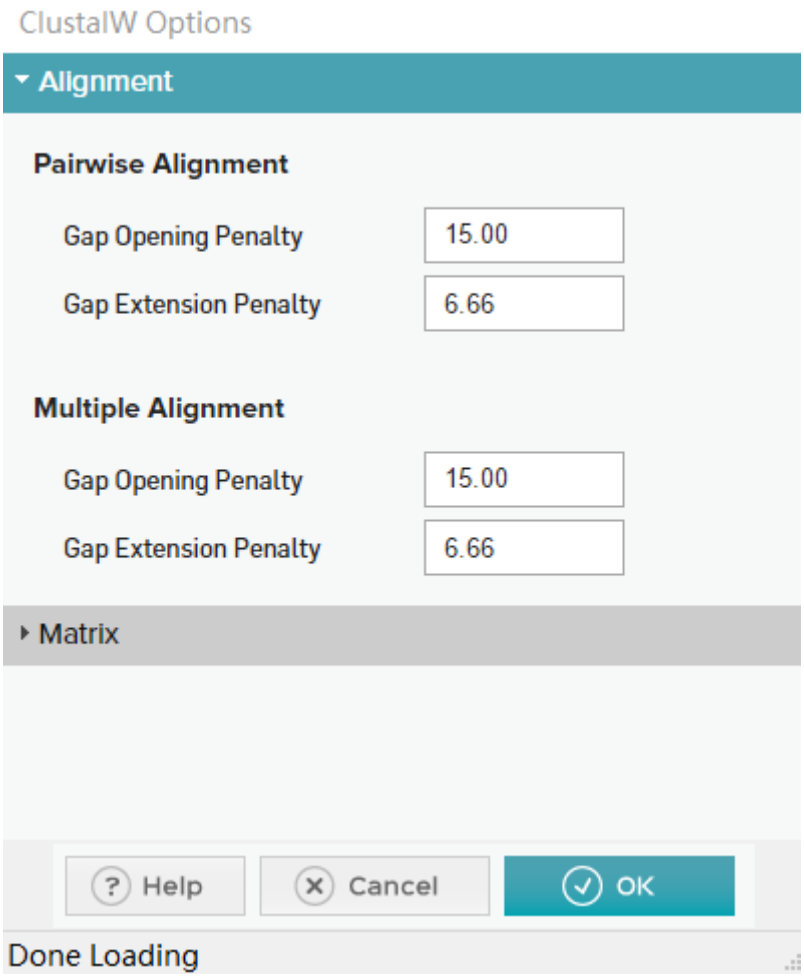


Fig. 3.5 ClustalW options

4. Phylogenetic analysis

   a) When alignment is completed, DNA sequences in the alignment explorer will be updated by the aligned ones.

   b) Then go to 'Data' → 'Phylogenetic Analysis'. A confirmation box will appear to ask if these sequences are protein-coding nucleotide sequence data or not. Please select 'Yes' or 'No' according to the property of the sequence target. In our example, these sequences can encode glycoproteins, so we select 'Yes' (**Fig. 3.6**).

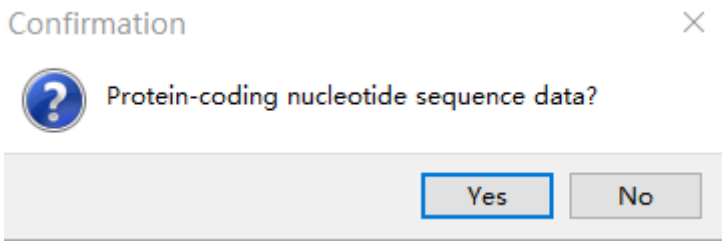


Fig. 3.6 Confirmation box to specify if sequences are protein-coding or not.

5. Select the NJ method

a) Return to the MEGA® main page.

b) Click the ‘PHYLOGENY’ icon (**Fig. 3.7**) and choose ‘Construct/Test Neighbor-Joining Tree’.

c) Then click ‘Yes’ to use the active file.

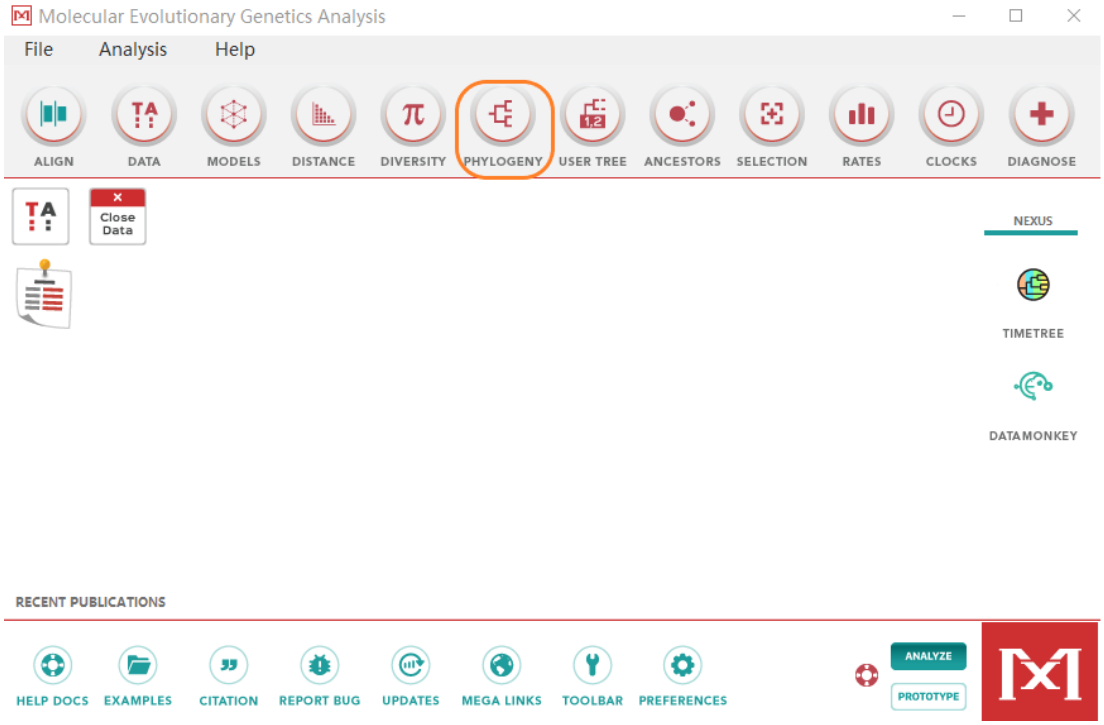


Fig. 3.7 Click the ‘PHYLOGENY’ icon to activate phylogenetic analysis.

6. Analysis preferences

a) After this, an Analysis Preferences box will pop up (**Fig. 3.8**).

b) Here we just need to specify ‘Bootstrap method’ for ‘Test of Phylogeny’, input 1000 for ‘No. of Bootstrap Replications’ and use default parameters for the other options.

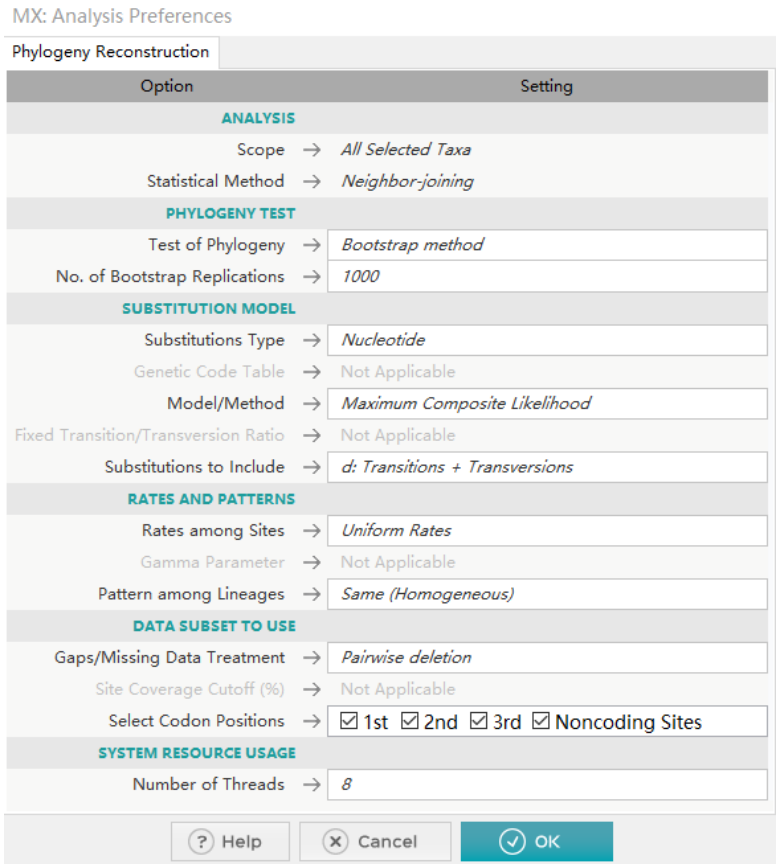


Fig. 3.8 Analysis preferences for NJ tree construction.

7. Phylogenetic tree

   a) After a while (the running time will depend on the data size and computer performance), a Tree Explorer will appear to show the generated phylogenetic tree (**Fig. 3.9**).

   b) The phylogenetic tree can be saved as image file by clicking ‘Image’ and selecting a suitable format. Usually, we save the tree file as an image in PDF or TIFF format, which can be used for publication.

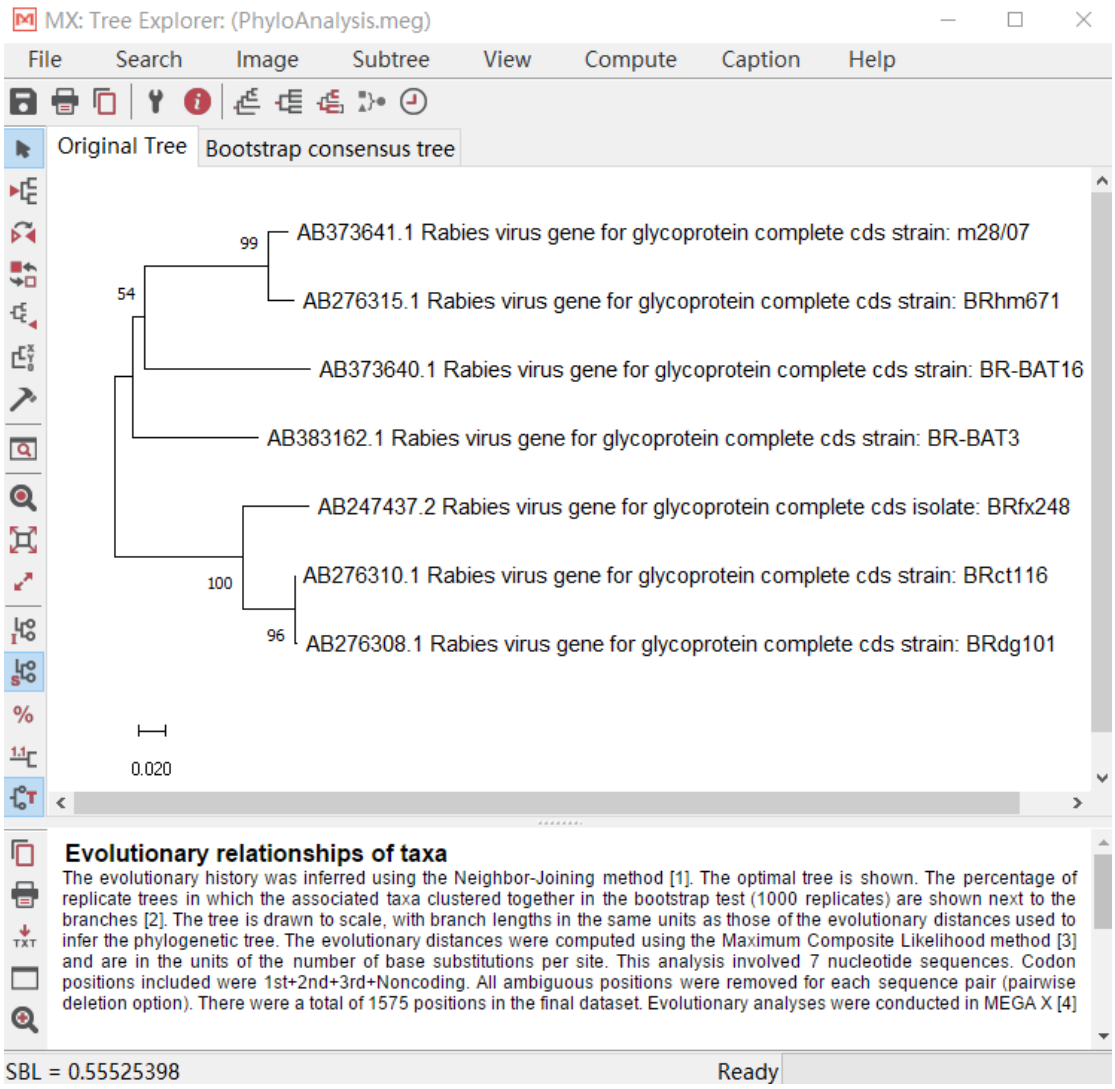


Fig. 3.9 NJ tree generated for the seven sequences.

8. Adding brackets to the phylogenetic tree

   a) Select the root-branch of a subtree, and click the hammer icon.

   b) In the Subtree Drawing Options, input a group name in the box for Name/Caption in the Property unit, and click 'OK' (**Fig. 3.10**)

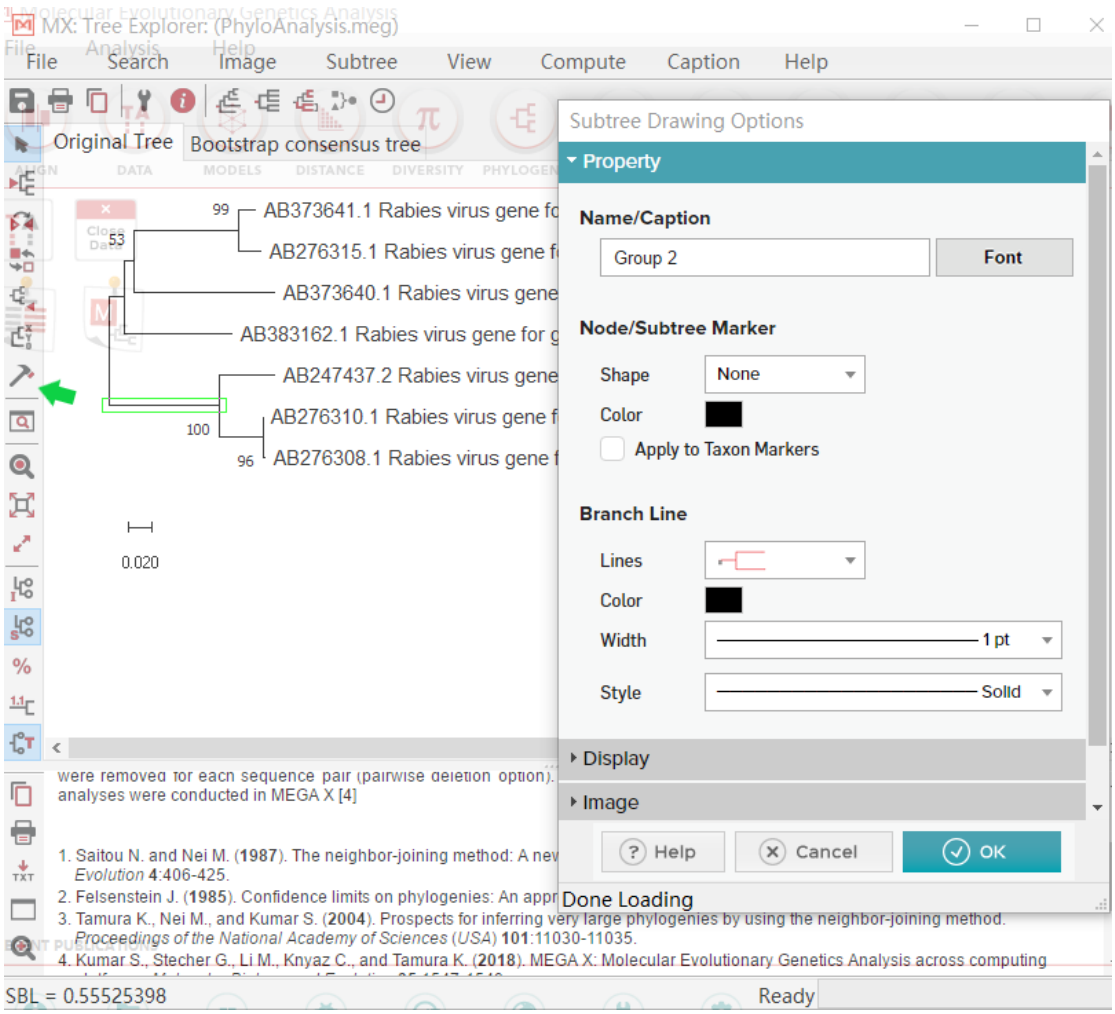


Fig. 3.10 Adding bracket to a subtree group.

9. Interpreting the phylogenetic tree

a) In the example, these RABV strains were clustered into Group 1 and Group 2 (**Fig. 3.11**). Strains BRct116 and BRdg101 are the two most closely-related viruses, since their phylogenetic branches are very short; this can be validated by their nucleotide identity (99.7%, see above result in the manual for BioEdit®). Background information also showed that BRct116 and BRdg101 were both detected in 2006, from cat and dog, respectively. Hence, they might have the same transmission source. Besides, m28/07 and BRhm671 are also closely related strains, sharing 96.5% nucleotide identity. BR-BAT16 and BR-BAT3 are two viruses distantly related to the other strains, since they have very long phylogenetic branches and both shared ≤ 83.1% nucleotide identities with the other viruses; this could be ascribed to their origin in bats.

b) Numbers next to the branches are the percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates). A higher number indicates more robust phylogenetic clustering. The bar below

the tree is the base substitution rate per site that is used to quantify the genetic distance between two sequences (**Fig. 3.11**).

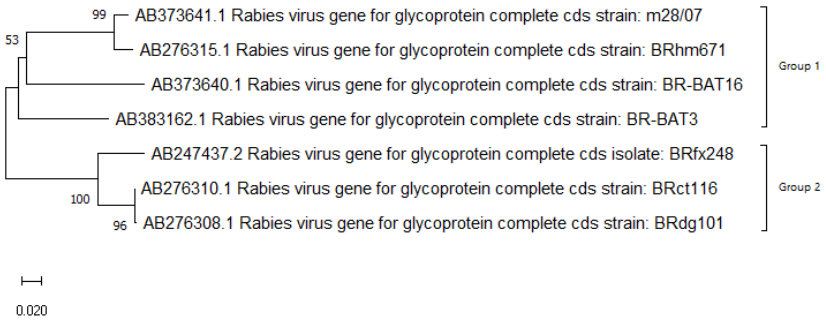


Fig. 3.11 The seven RABV strains were clustered into two groups.