

# Multi-party genome analysis platform for hCoV-19(SARS-CoV-2) sequence -- Instruction

## Introduction

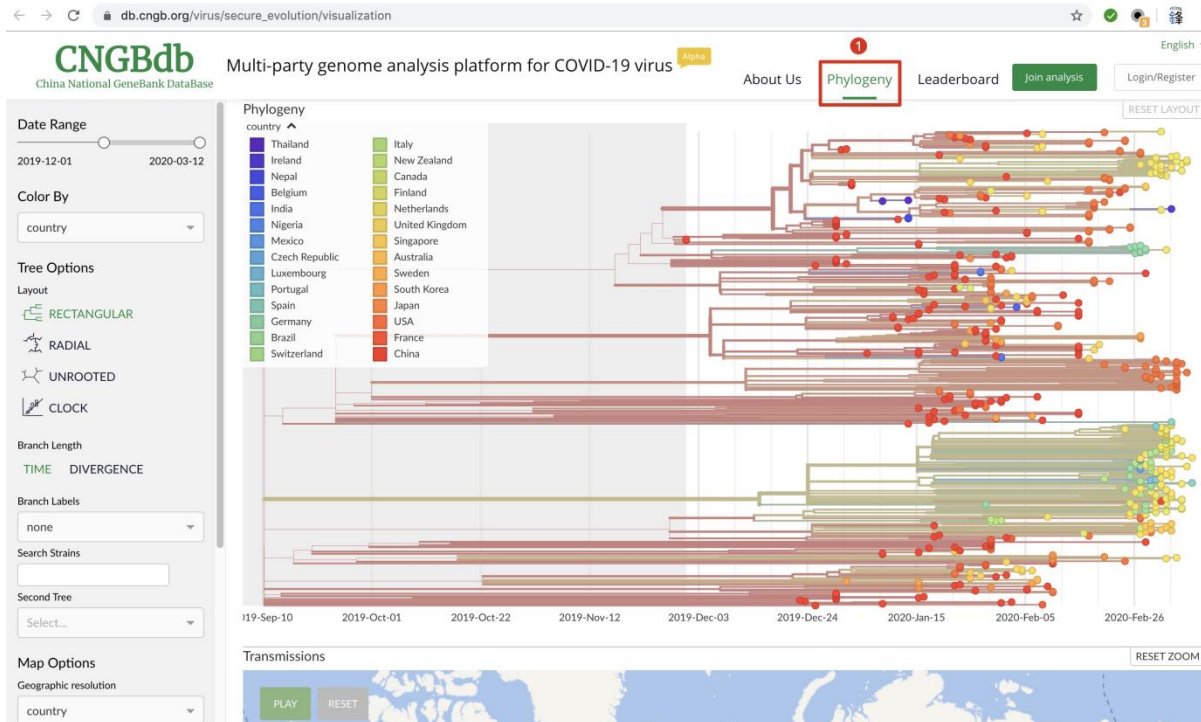
At the beginning of 2020, COVID-19 pandemic caused by the SARS-CoV-2 (formerly known as 2019-nCov) virus broke out across the world. Currently, potential vaccines and drugs are still under development. It is particularly important to share the genome sequences of the coronavirus to provides scientific evidence for disease testing, control, monitoring and drug development. To promote data sharing and track the real-time virus evolution status, China National GeneBank and BGI Blockchain team jointly developed and launched this platform for researchers to share and perform joint analysis of the coronavirus genome online.

## Secure Multi-Party Computation and Blockchain

By using the Multi-party genome analysis tools, all parties can perform joint data analysis and share the results without publishing the original data. The platform uses the alliance chain to verify the dataset of each party. The data usage history and analysis history will be stored and verified on the blockchain. The verification step ensures that your dataset you are always associated with you. Even if you choose to make your data available for others to use in analysis, all related analysis and history can be traced back to the initial data you provided. The data verification records on blockchain ensures that they cannot be tampered with, thus provides reliable tracking of the data flow. Additionally, once the data verification is on chain, the data provided by the user is guaranteed to be consistent with the content of multi-party analysis protocol.

## Our platform provides the following functions

1. **Real-time display of the evolution of currently available virus sequences from CNGBdb, GISAID, NCBI and other databases**



Display the evolutionary tree based on the currently available public data. We obtain these sequences data from various public databases (such as CNGBdb, GISAID, NCBI etc.). For the tree displayed here, we only include virus whole genome sequences from human hosts.

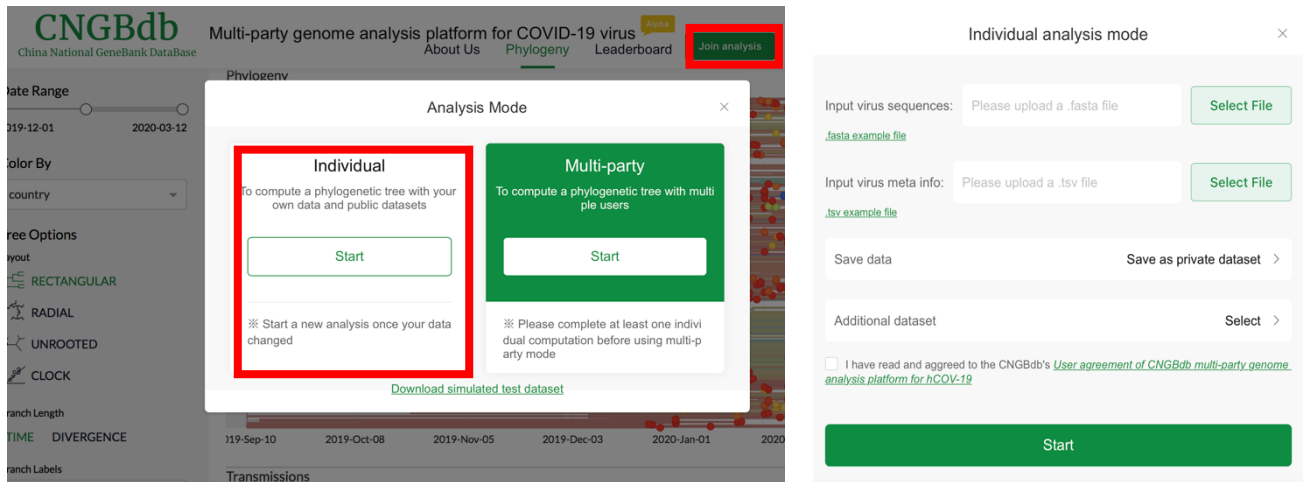
### The results displayed include:

- Virus genome evolutionary relationships
- Sequence mutation information
- Geographic location of samples
- Sampling time
- Data hash information on blockchain

### You may adjust the display settings on the left side panel:

- Change the tip color according to sample geographic location, sampling date etc.
- Set the phylogeny layout
- Set the branch length to sampling time or divergence based on sequence mutations
- Show virus evolution animation based on sampling time and inferred phylogeny

## 2. Perform phylogenetic analysis with your own data – individual mode



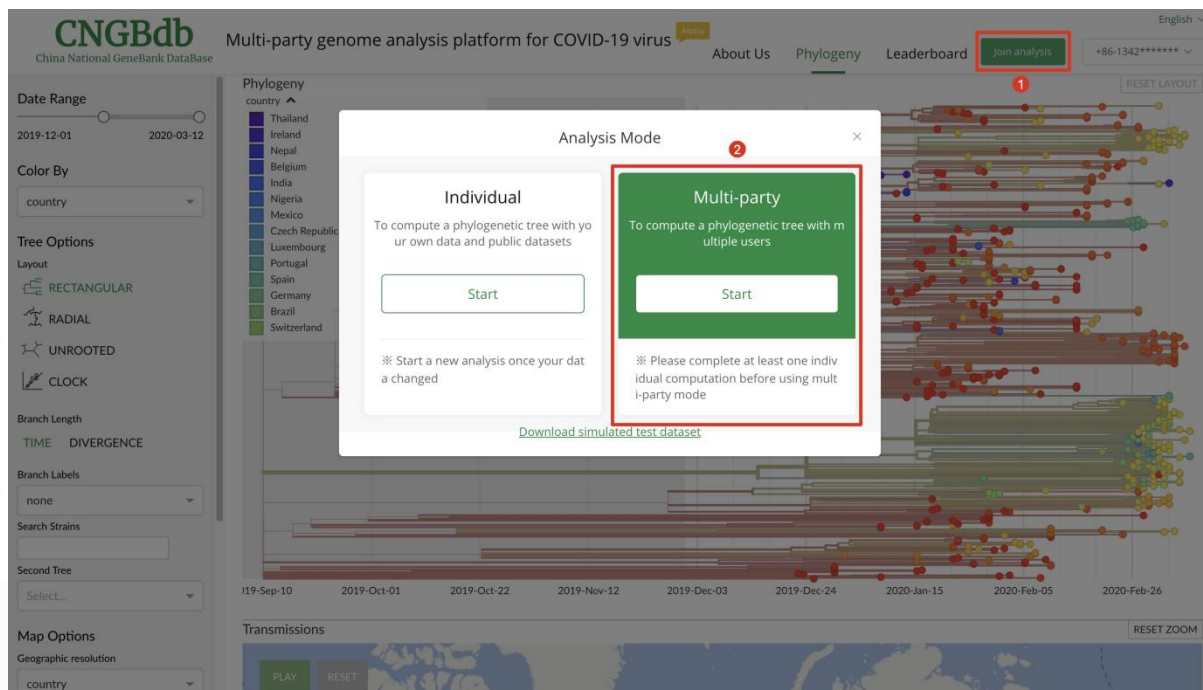
In this mode, user can analyze the relative position of his/her private virus sequences on the current public phylogenetic tree. Registered users can click "Join analysis"-> "Individual mode", and combine user's private data and public data stored on the platform to calculate the relative position of the private virus sequence on the current public evolution tree.

### Steps:

1. Upload user's input files, including a .fasta file for sample sequences and a .tsv file for sample meta information. Example input files can be downloaded on "Individual analysis mode" page.
2. Select how you want to save the data:
  - Save as private dataset  
Your data will be saved on the platform as your personal dataset. Only you can access or use it for analysis. Authorization is needed to use personal data in multi-party analysis. Be aware that your result (tree) will always be public viewable.
  - Save as public computable dataset  
Your data will be saved on the platform as a public computable dataset. Other users cannot directly see or download your data but can add your dataset as an additional dataset in their analysis.

- Do not save  
Your data won't be saved on the platform. It'll only be used in this particular analysis and you need to upload it again if want to use it in multi-party analysis mode later. Be aware that your result (tree) will always be public viewable.
3. Select if you want to add additional dataset to your analysis, additional dataset includes:
- Public datasets obtained from CNCBdb, GISAID, NCBI etc.
  - Public computable datasets contributed by other users.

### 3. Perform phylogenetic analysis with other users – multi-party mode



The screenshot shows the CNCBdb Multi-party genome analysis platform for COVID-19 virus. The interface includes a 'Join analysis' button (1) and an 'Analysis Mode' dialog box (2) with 'Individual' and 'Multi-party' options. The background displays a phylogenetic tree and a transmission map.

Multi-party analysis mode - Initiate protocol

Participant ID	cngbdb_0790493749
History Dataset	20200310113859096_cngbdb_0790493749_2 >
Additional dataset	Select >
My Dataset	20200330185849_cngbid_a48396318403d_1 >

[Add participant](#)

I have read and agreed to the CNCBdb's [User agreement of CNCBdb multi-party genome analysis platform for hCOV-19](#)

[Save draft](#)

[Release protocol](#)

Multi-party analysis mode - sign protocol

Data to be used	
Participant 1 user ID:	cngbdb_0790493749
Participant dataset	20200310113859096_cngbdb_0790493749_2 <span>Waiting</span>
Additional dataset	None

My Dataset	
My saved data	20200330185849_cngbid_a48396318403d_1 <span>Waiting</span>
Upload new .fasta file	<a href="#">Select File &gt;</a>
<a href="#">.fasta example file</a>	
Upload new .tsv file	<a href="#">Select File &gt;</a>
<a href="#">.tsv example file</a>	

I have read and agreed to the CNCBdb's [User agreement of CNCBdb multi-party genome analysis platform for hCOV-19](#)

[Verify data and sign](#)

If all parties signed the protocol and passed the data verification, protocol will be executed automatically

In this mode, multiple users can perform joint phylogenetic analysis with their own data. Registered users can initiate joint analysis through "Join analysis" -> "Multi-Party Mode". All parties can perform joint data analysis and share the results without revealing the original data. The platform uses the alliance chain to verify the dataset of each party.

A user may initiate a multi-party analysis protocol and invite others or join an analysis other user initiated as a participator.

### Steps:

1. Initiate a protocol

The multi-party computing protocol specifies which user should join the analysis, and the data sets that each participant should use (that is, the data sets used in an individual analysis performed by each participant before).

2. Protocol signing phase

Since the platform performs certification for each analysis and data set, the data set provided by the participant during the multi-party analysis can be verified by calling the blockchain certification information to verify that if it is consistent with the data set specified in the protocol. The successful signing of the protocol needs to meet two conditions: the participants acknowledge and agree to the content of the multi-party analysis protocol, and the participants provide a legitimate data set.

3. Protocol execution

After all parties involved have successfully signed, the protocol will automatically enter the calculation execution step. If any participant fails to sign, the agreement will not be executed.