

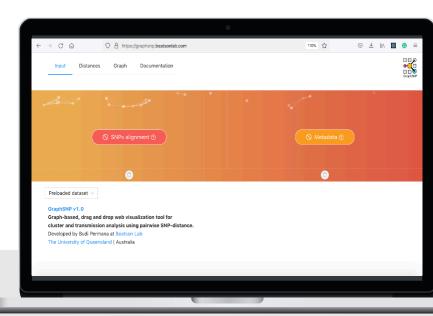
## GraphSNP USER MALL

## **Budi Permana**

v.2022.05.01

## Using GraphSNP

GraphSNP is a single page application (SPA) visualisation tool that runs on the browser. Users can visualise and explore data by loading their input files or setting up multiple projects (available on offline use only) for multiple input datasets.



https://graphsnp.beatsonlab.com/



(npm install serve -g)

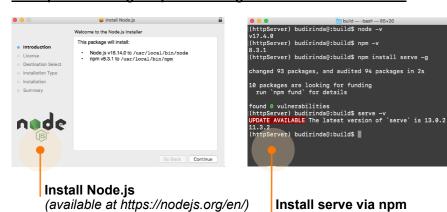
## Use it online

GraphSNP is deployed in https:// graphsnp.beatsonlab.com for online use. Users can visit the web page using modern browsers (e.g., Google Chrome, Firefox, Microsoft Edge, and Safari), drag and drop the input files, and instantly perform interactive data visualization and analysis.

## Use it offline

Users also can use GrapSNP offline by serving it through a local HTTP server. GraphSNP SPA can be downloaded from https://github.com/nalarbp/graphsnp/build/.

## Example of serving GraphSNP using HTTP-server "serve" tool



budirinda@:build\$ serve AVAILABLE The latest version of `serve` is 13.0.2 - Local: http://localhost:5000 - On Your Network: http://10.89.122.236:5000

> **Run the HTTP server** (serve .)

## Input Files

## SNPs alignment

A text file containing a minimum of two equal lengths of fasta-formatted non-gap nucleotide sequences (N can be used to represent ambiguous nucleotides).

## Example SNPs alignment input (sample.fasta)

ATTGCAGCTATGTTGACGATGAC ATTGCAGCTAGACAGACGATGAC CGAATGAGCCTGTTGTAGATGAC

ATTGCAGCTAGACACACGATGAC CGAGCAGCTATGTTGACCCACGT

ATTGCAGCTAGACAGACGATGAC

1	A	T'	Γ	G(	CA	G	C	Т	Α	Т	G	Т	Т	G	Α	C	G	A	Т	G	Α	C
2	A	T'	Γ	G	CA	G	C	Т	A	G	Α	C	A	G	A	C	G	A	Т	G	A	C
		G																				
4	A	T'	Γ	G	CA	G	C	Т	A	G	A	C	A	G	A	C	G	A	Т	G	A	C
5	A	T'	Γ	G	CA	G	C	Т	A	G	A	C	A	C	A	C	G	A	Т	G	A	C
6	C	G	Δ	G	קי	G	C	Т	Δ	Т	G	Т	т	G	Δ	C	C	$\boldsymbol{C}$	Δ	C	G	т

Sample ID in fasta header

## **Example of pairwise SNP distances matrix** (sample matrix.csv)

dist	1	2	3	4	5	6
1	0	4	12	4	5	9
2	4	0	16	0	1	13
3	12	16	0	16	17	15
4	4	0	16	0	1	13
5	5	1	17	1	0	14
6	9	13	15	13	14	0

Mandatory column for

## **Matrix in CSV format**

dist, 1, 2, 3, 4, 5, 6 1,0,4,12,4,5,9 2,4,0,16,0,1,13 3,12,16,0,16,17,15 4,4,0,16,0,1,13 5,5,1,17,1,0,14 6,9,13,15,13,14,0

User can also input the pairwise distances matrix

Pairwise distances matrix

instead of SNP alignment. The symmetric matrix should be written in comma-separated value (CSV) format.

## Metadata

Mandatory

A table contains information about the isolates or sample, written in CSV format. Critical requirements including: mandatory headers, no duplicated records in column sample\_id. Column collection\_day is required for transmission analysis.

Any additional column

column	transmission analys	ıs	<u> </u>				<del>'</del>
sample_id	collection_day	Location	Source	Clade	Gene-A	Source:color	Gene-A:color
1	1	room A	clinical	Α	present	#FF8076	Black
2	2	room B	clinical	Α	present	#FF8076	Black
3	3	room C	clinical	Α	present	#FF8076	Black
4	3	room A	environmental	Α	absent	#53DE22	White
5	4	room B	environmental	Α	absent	#53DE22	White
6	5	room C	environmental	Α	absent	#53DE22	White

Columns to set the color

## Page navigation

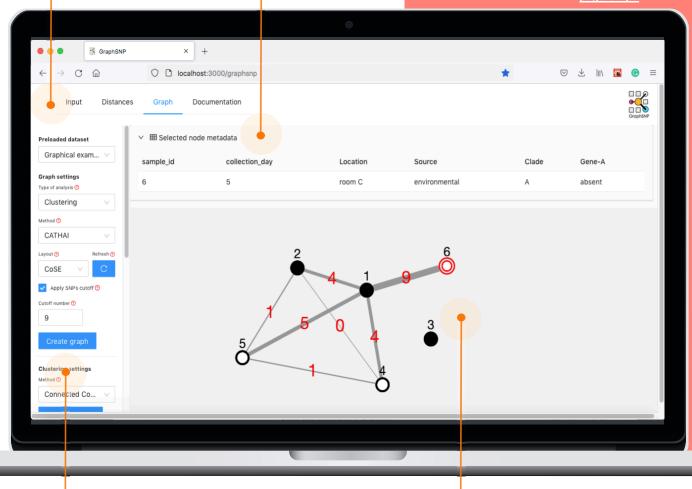
Navigation menu to let you jump between pages: Input, Distances, Graph, and Documentation.

## Metadata table

Let you display metadata associated with selected node(s).

## Main nterface

page Graph



## Sidebar settings

A sidebar menu provides you a control to adjust the visualisation.

## **Graph visualisation window**

A window container where the interactive graph is being rendered.

## page Input

# × +

## page Distances



## Input placeholder

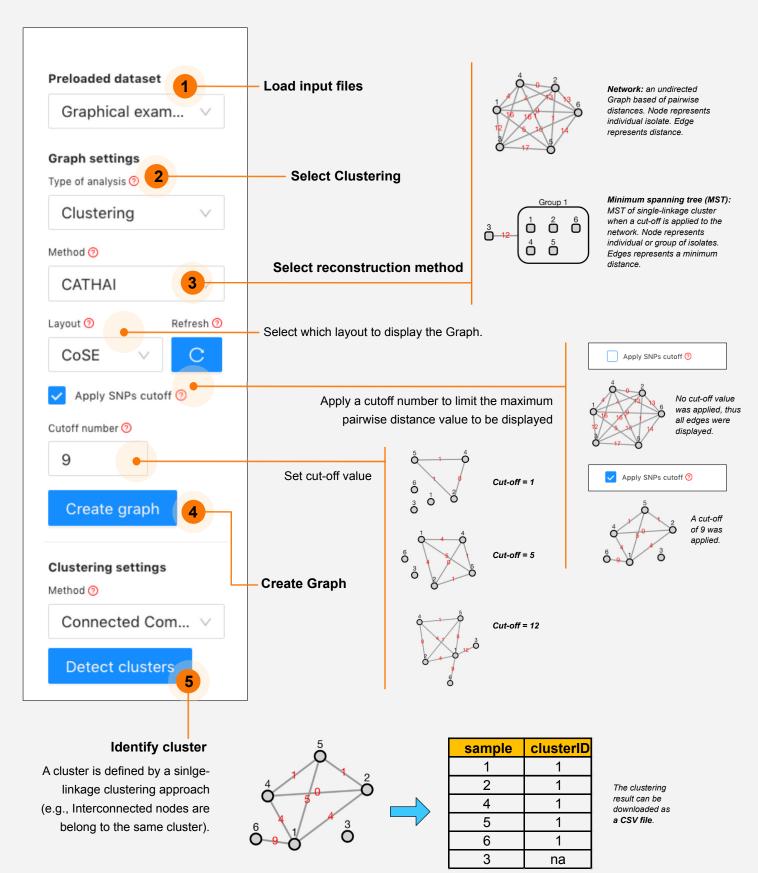
Drag and drop your input files here.

## **Chart visualisation window**

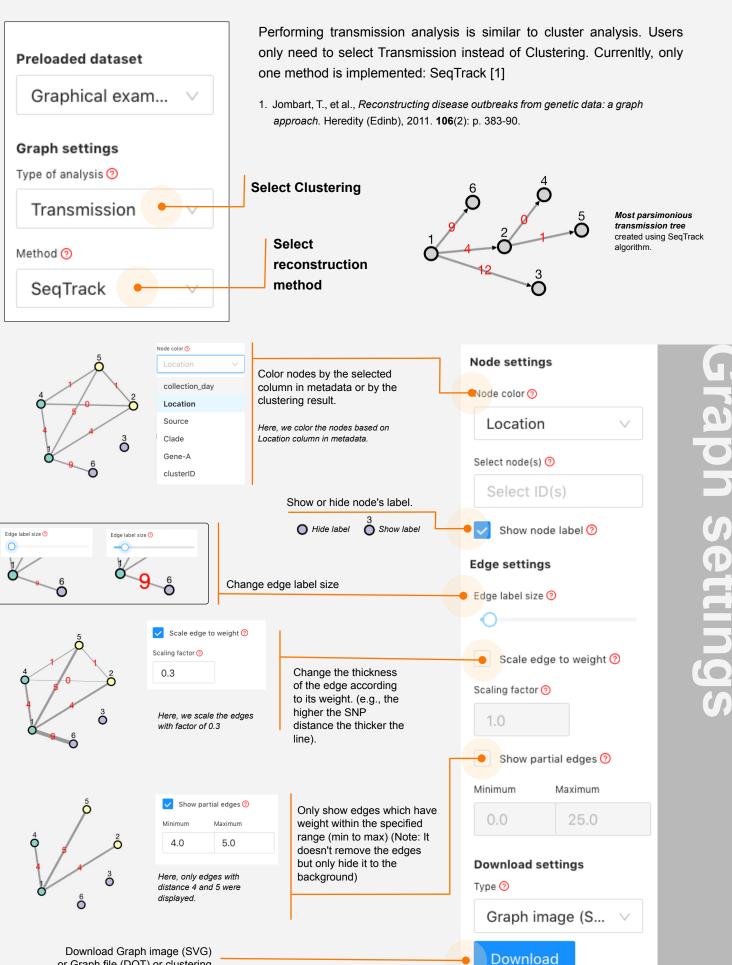
A container where charts showing pairwise distances, like bar chart, is being rendered.

## Cluster analysis

Users can perform cluster analysis and visualization by five simple steps: Loading the input files, select clustering as the type of analysis (another type is transmission analysis), select the clustering method, create the Graph and detect cluster from the Graph.



## Transmission analysis



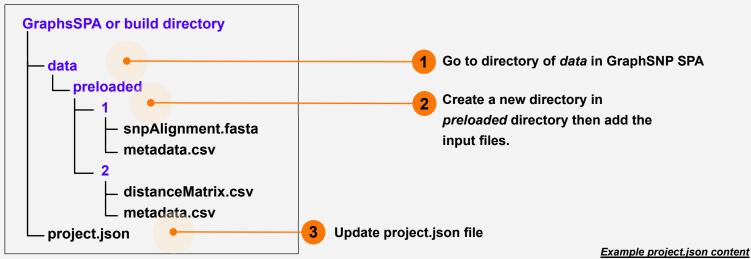
or Graph file (DOT) or clustering

result (CSV)

## Setting up preloaded dataset

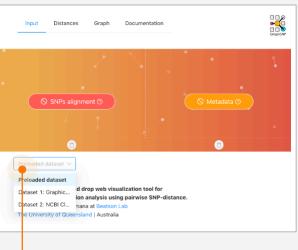
When users use GraphSNP offline, they can set up multiple preloaded datasets. This feature allows users to 'permanently' link their input files to GraphSNP, avoiding the need to re-inputting their input files every time the browser refreshed.

### Example of directory tree of GraphSNP preloaded datasets



Add the dataset ID and input files path to project.json and save the file.

Datasets is listed in GraphSNP input page



Klick the preloaded dataset dropdown button and select dataset of interest and GraphSNP will automatically load the input files.

```
"projects": [
    "id": "1",
    "name": "Dataset 1: Graphical example",
    "matrixOrAlignment": "alignment",
    "snpDistance": "./data/preloaded/1/snpAlignment.fasta",
    "metadata": "./data/preloaded/1/metadata.csv"
   },
    "id": "2",
    "name": "Dataset 2: NCBI Cluster of VREfm ST78",
    "matrixOrAlignment": "matrix",
    "snpDistance": "./data/preloaded/2/distanceMatrix.csv",
    "metadata": "./data/preloaded/2/metadata.csv"
  ],
  "description": "This JSON file describes preloaded datasets to be rendered in
the landing page. The path of these files must be written with directory 'public' as
the root (e.g. ./data/ means 'data' is inside directory 'public'
```



for reading this manual









Thanks to all awesome web frameworks and libraries run on the background, GraphSNP is now up and running and available worldwide. The following are some of the core libraries used by GraphSNP:

react d3 antd cytoscape cytoscape-svg redux react-color @nivo lodash moment moment-range

