

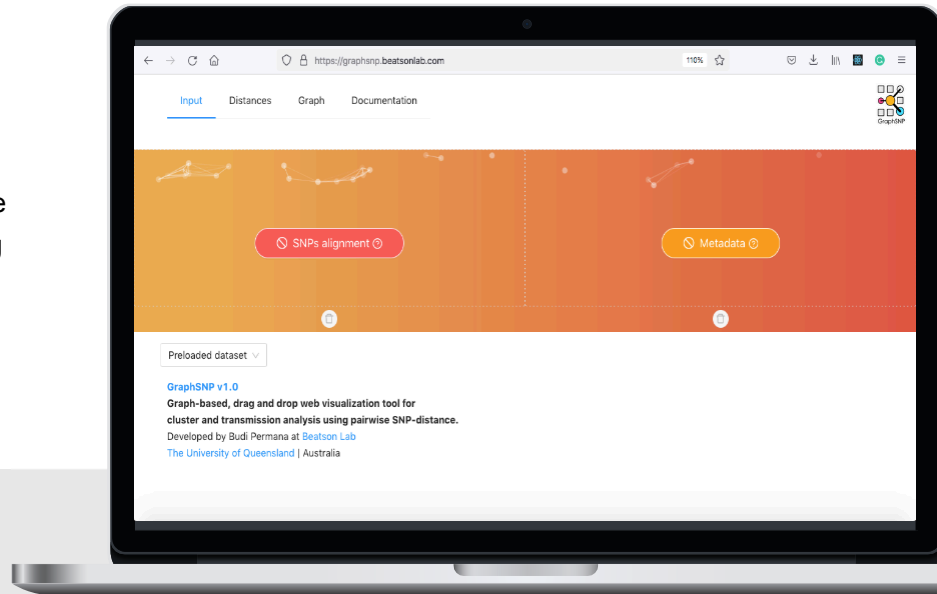
GraphSNP USER MANUAL

Budi Permana

v.2022.04.01

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.



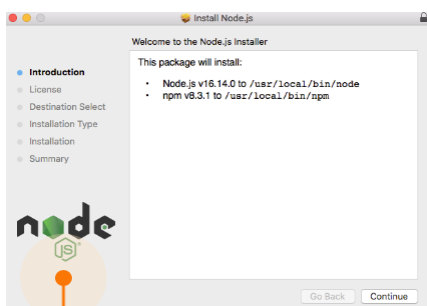
• Use it online

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.

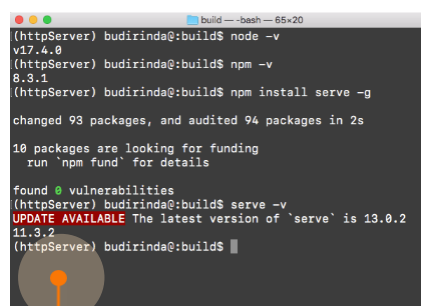
• Use it offline

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/>.

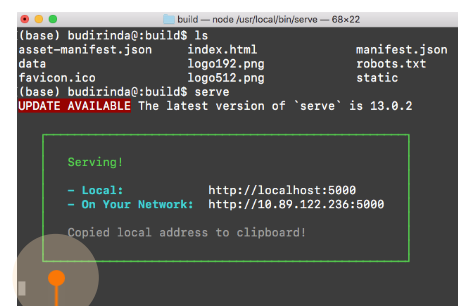
Example of serving GraphSNP using HTTP-server “serve” tool



Install Node.js
(available at <https://nodejs.org/en/>)



Install serve via npm
(`npm install serve -g`)



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*)

```
>1
ATTGCAGCTATGTTGACGATGAC
>2
ATTGCAGCTAGACAGACGATGAC
>3
CGAATGAGCCTGTTGTAGATGAC
>4
ATTGCAGCTAGACAGACGATGAC
>5
ATTGCAGCTAGACACACGATGAC
>6
CGAGCAGCTATGTTGACCCACGT
```

1	A	T	T	G	C	A	G	C	T	A	T	G	T	T	G	A	C	G	A	T	G	A	C
2	A	T	T	G	C	A	G	C	T	A	G	A	C	A	G	A	C	G	A	T	G	A	C
3	C	G	A	A	T	G	A	G	C	C	T	G	T	T	G	T	A	G	A	T	G	A	C
4	A	T	T	G	C	A	G	C	T	A	G	A	C	A	G	A	C	G	A	T	G	A	C
5	A	T	T	G	C	A	G	C	T	A	G	A	C	A	C	A	C	G	A	T	G	A	C
6	C	G	A	G	C	A	G	C	T	A	T	G	T	T	G	A	C	C	C	A	C	G	T

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

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

• Metadata

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.

Mandatory column	Mandatory column for transmission analysis	Any additional column				Columns to set the color	
sample_id	collection_day	Location	Source	Clade	Gene-A	Source:color	Gene-A:color
1	1	room A	clinical	A	present	#FF8076	Black
2	2	room B	clinical	A	present	#FF8076	Black
3	3	room C	clinical	A	present	#FF8076	Black
4	3	room A	environmental	A	absent	#53DE22	White
5	4	room B	environmental	A	absent	#53DE22	White
6	5	room C	environmental	A	absent	#53DE22	White

Main interface

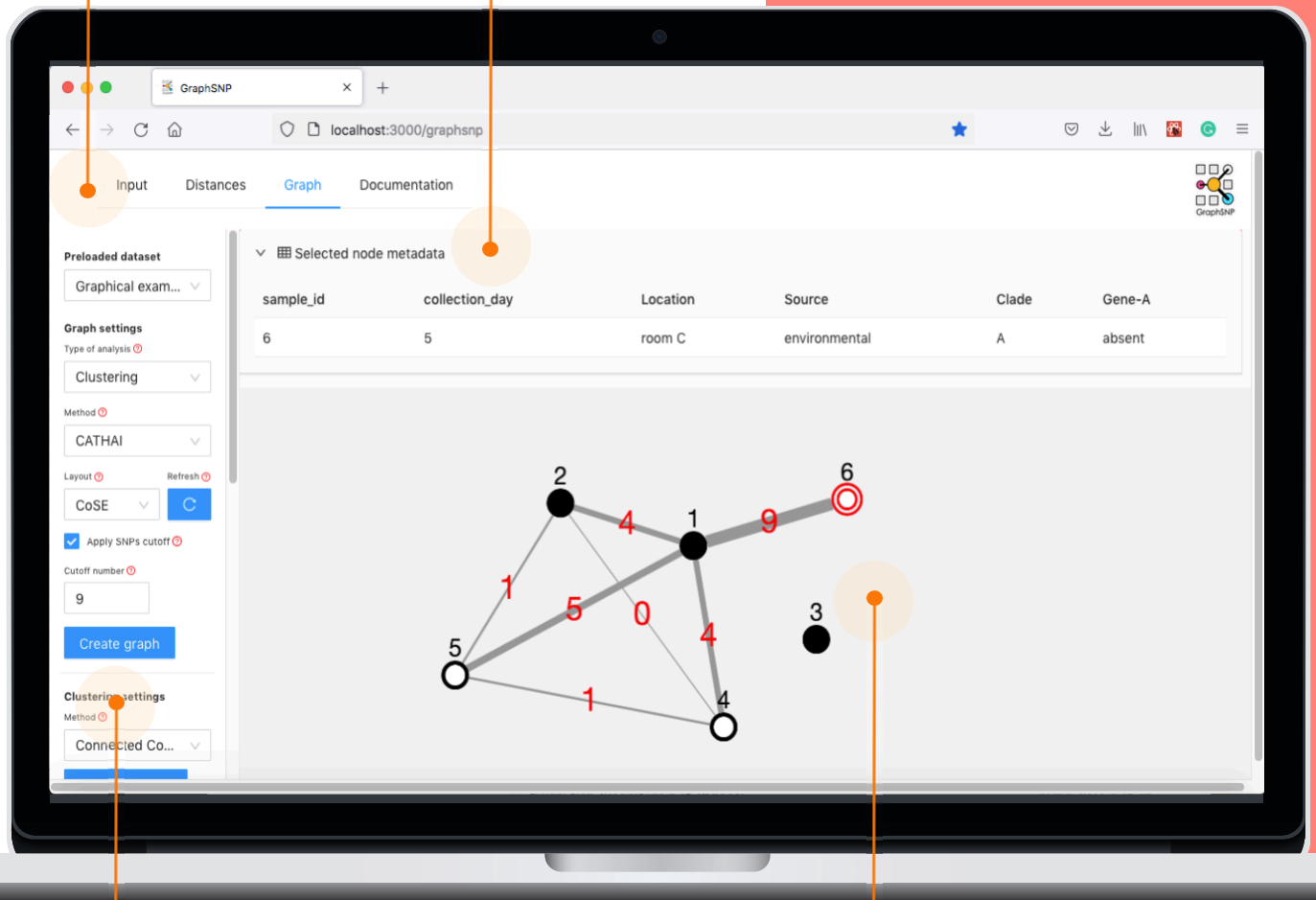
Page navigation

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

Metadata table

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

page *Graph*



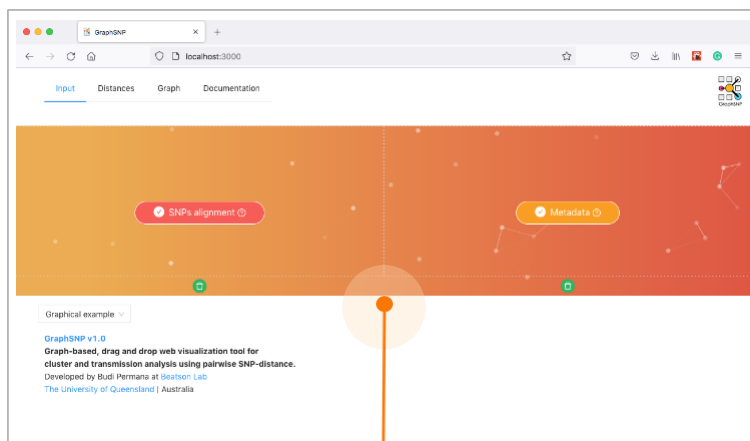
Sidebar settings

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*



Input placeholder

Drag and drop your input files here.

page *Distances*

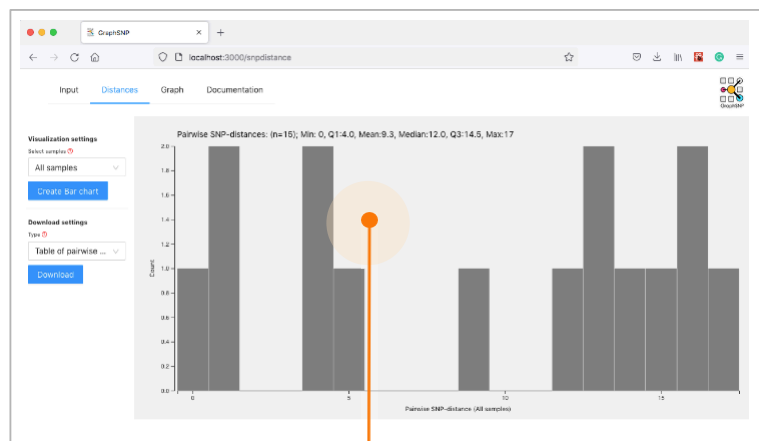


Chart visualisation window

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.

Preloaded dataset

Graphical exam... ▾

Graph settings

Type of analysis ▾

Clustering ▾

Method ▾

CATHAI

Layout ▾

CoSE ▾

☒ Apply SNPs cutoff

Cutoff number

9

Create graph

Clustering settings

Method ▾

Connected Com... ▾

Detect clusters

Load input files

Select Clustering

Select reconstruction method

Select which layout to display the Graph.

Apply a cutoff number to limit the maximum pairwise distance value to be displayed

Set cut-off value

Create Graph

Network: an undirected Graph based of pairwise distances. Node represents individual isolate. Edge represents distance.

Minimum spanning tree (MST): MST of single-linkage cluster when a cut-off is applied to the network. Node represents individual or group of isolates. Edges represents a minimum distance.

☐ Apply SNPs cutoff

No cut-off value was applied, thus all edges were displayed.

☒ Apply SNPs cutoff

A cut-off of 9 was applied.

Cut-off = 1

Cut-off = 5

Cut-off = 12

Identify cluster

A cluster is defined by a single-linkage clustering approach (e.g., Interconnected nodes are belong to the same cluster).

sample	clusterID
1	1
2	1
4	1
5	1
6	1
3	na

The clustering result can be downloaded as a CSV file.

Transmission analysis

Preloaded dataset

Graphical exam... ▾

Graph settings

Type of analysis ?

Transmission ▾

Method ?

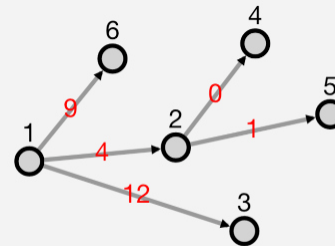
SeqTrack ▾

Performing transmission analysis is similar to cluster analysis. Users only need to select Transmission instead of Clustering. Currently, only one method is implemented: SeqTrack [1]

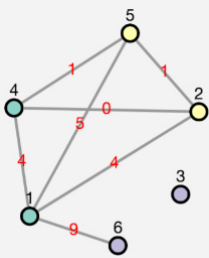
1. Jombart, T., et al., *Reconstructing disease outbreaks from genetic data: a graph approach*. Heredity (Edinb), 2011. **106**(2): p. 383-90.

Select Clustering

Select reconstruction method



Most parsimonious transmission tree created using SeqTrack algorithm.



Node color ?

Location ▾

collection_day

Location

Source

Clade

Gene-A

clusterID

Color nodes by the selected column in metadata or by the clustering result.

Here, we color the nodes based on Location column in metadata.

Node settings

Node color ?

Location ▾

Select node(s) ?

Select ID(s)

Show or hide node's label.

☐ Hide label ☒ Show label

Edge settings

Edge label size ?

Scale edge to weight ?

Scaling factor ?

1.0

Show partial edges ?

Minimum Maximum

0.0

25.0

Download settings

Type ?

Graph image (S... ▾

Download

Change edge label size

Change the thickness of the edge according to its weight. (e.g., the higher the SNP distance the thicker the line).

☒ Scale edge to weight ?

Scaling factor ?

0.3

Here, we scale the edges with factor of 0.3

☒ Show partial edges ?

Minimum Maximum

4.0 5.0

Here, only edges with distance 4 and 5 were displayed.

Only show edges which have weight within the specified range (min to max) (Note: It doesn't remove the edges but only hide it to the background)

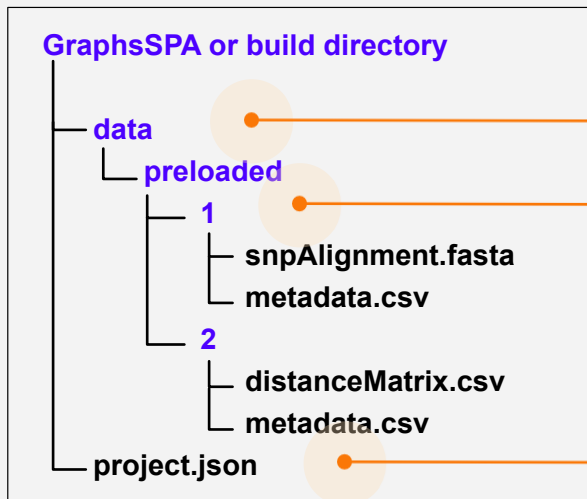
Download Graph image (SVG) or Graph file (DOT) or clustering result (CSV)

Graph settings

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



1 Go to directory of *data* in GraphSNP SPA

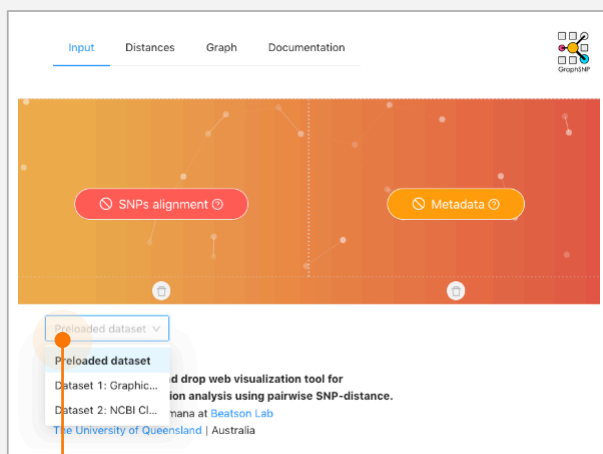
2 Create a new directory in *preloaded* directory then add the input files.

3 Update *project.json* file

Example *project.json* content

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

4 Datasets is listed in GraphSNP input page



```
{
  "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'"
}
```

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

THANK YOU

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
...

