

# Web-gLV user manual

## 1. Data input:

Web-gLV can perform simulations starting two types of input sets:

- (a) Only a taxonomic abundance file which will be used to estimate parameters and generate reference plots for the observed trends for easy comparison. The reference values of initial starting point of simulation for the selected taxa set is also taken from one of the time point row of this abundance table.
- (b) A taxonomic abundance file, a user supplied growth rate file and inter-taxa interaction file.

File formats: All the input files needs to be **tab delimited** and in a format as shown below:

**Abundance file**

First column should contain the time-delta values ←

	Taxa1	Taxa2	Taxa3	Taxa4
0	30	3259	7454	813
1	408	4548	16240	5942
2	1405	5325	22250	27211
3	312	3172	10575	17139
4	754	5956	18852	22484
6	1322	5733	22060	25420
8	1301	2105	17997	21199
10	1931	3524	25516	25422
14	2064	3008	18681	18389
17	3734	5863	29824	30665
21	2338	3943	38292	36988
24	1555	2903	18553	25254
28	1835	2237	34647	33251

**Growth rate file**

Taxa	GrowthRate
Taxa1	-0.1186097278
Taxa2	3.198611059
Taxa3	1.5446019575
Taxa4	-1.7734226041

**Inter taxa interaction file**

Taxa	Taxa1	Taxa2	Taxa3	Taxa4
Taxa1	0.8.60E-006	-7.36E-006	-7.35E-006	2.10E-005
Taxa2	3.08E-005	-7.03E-005	-7.83E-005	6.80E-005
Taxa3	-3.35E-005	-7.91E-006	2.24E-005	-2.19E-005
Taxa4	4.91E-005	-7.44E-006	2.49E-005	-6.98E-006

An optional tab delimited metadata file corresponding to the time delta can also be uploaded specifying a RGB color in the format below:

**Metadata file**

TimeStart	TimeEnd	Metadata	Color
0	10	State1	255,255,0
10	28	State2	255,0,0

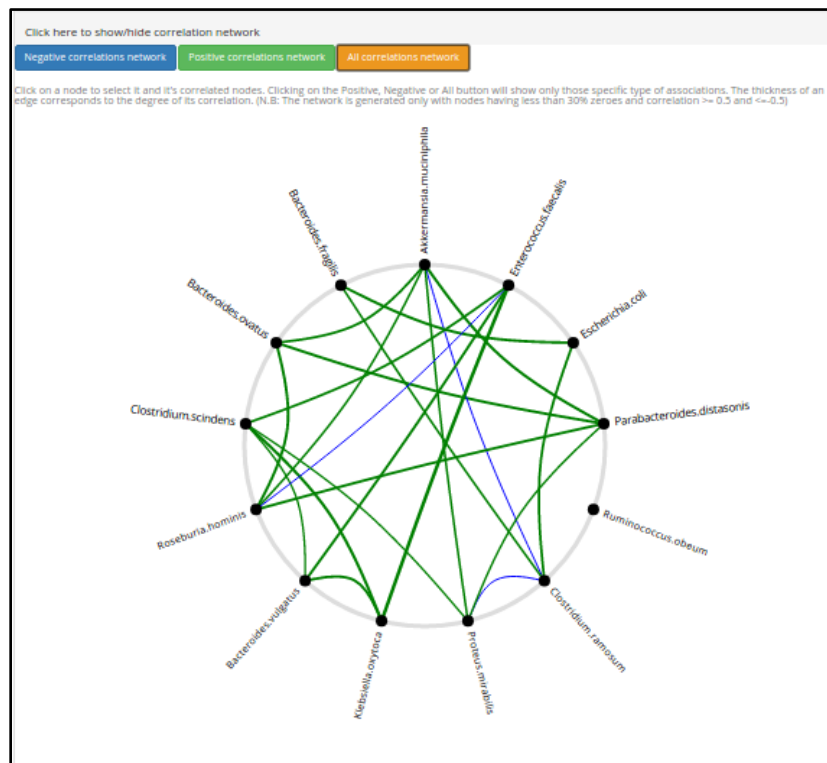
## 2. Running a simulation:

### Step 1: Select the taxa required for simulation from the input dataset

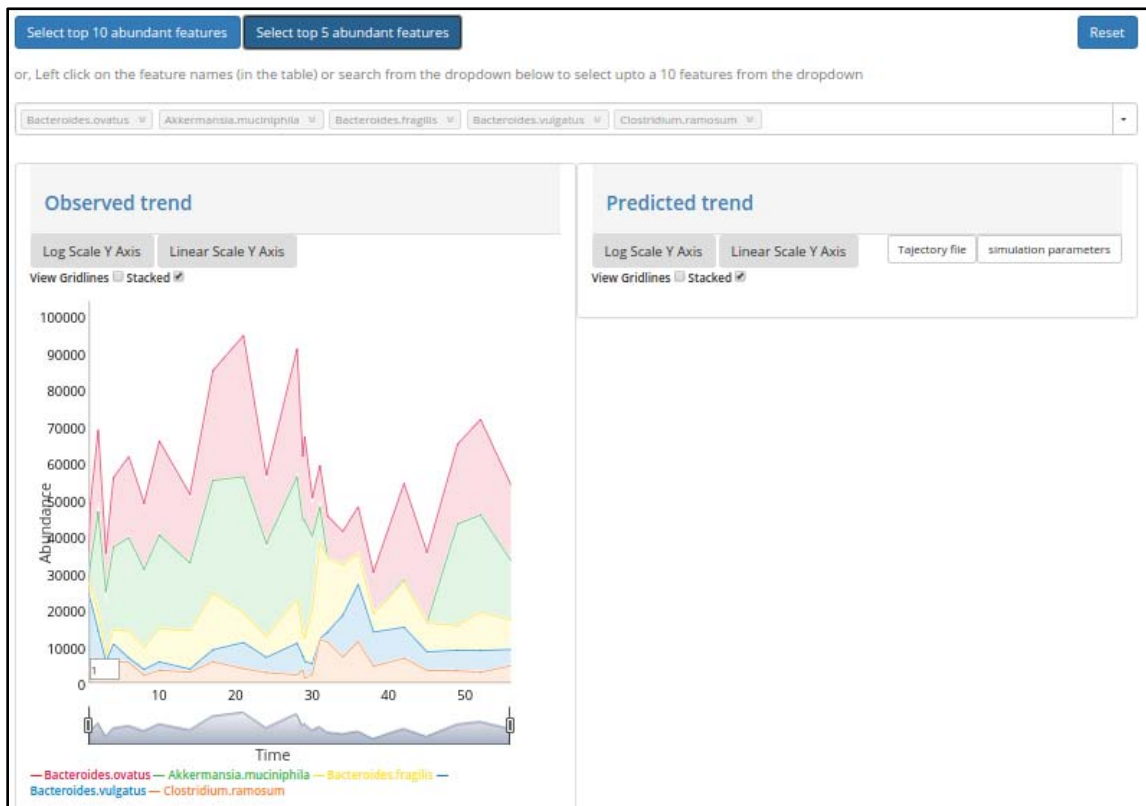
A summary of the input taxonomic abundance matrix is presented as an interactive table which can be used to sort and view the taxa abundances. By default, the table is sorted in descending order of taxa mean abundance. A taxon can be selected for simulation by clicking on its label from the table.

Data summary							
Name	Mean	Median	SD	% Zero	Trend plot	Box plot	
<input type="text" value="filter column..."/>							
Bacteroides.ovatus	19237.3846	18766.5	7811.05	0			
Akkermansia.muciniphila	17436.0385	20438	12608.12	0			
Bacteroides.fragilis	9176.6923	7910.5	5473.14	0			
Bacteroides.vulgatus	6285.8077	4637	5235.76	0			
Clostridium.ramosum	4862.5385	3764	2834.26	0			
Escherichia.coli	4025.6154	2866	2613.93	0			
Klebsiella.oxytoca	3226.7308	1490	5225.17	0			
Ruminococcus.obelum	1975	1809.5	1776.25	0			
Roseburia.hominis	1550.8846	1544	896.23	0			
Parabacteroides.distasonis	933.5769	891.5	832.24	7.69			

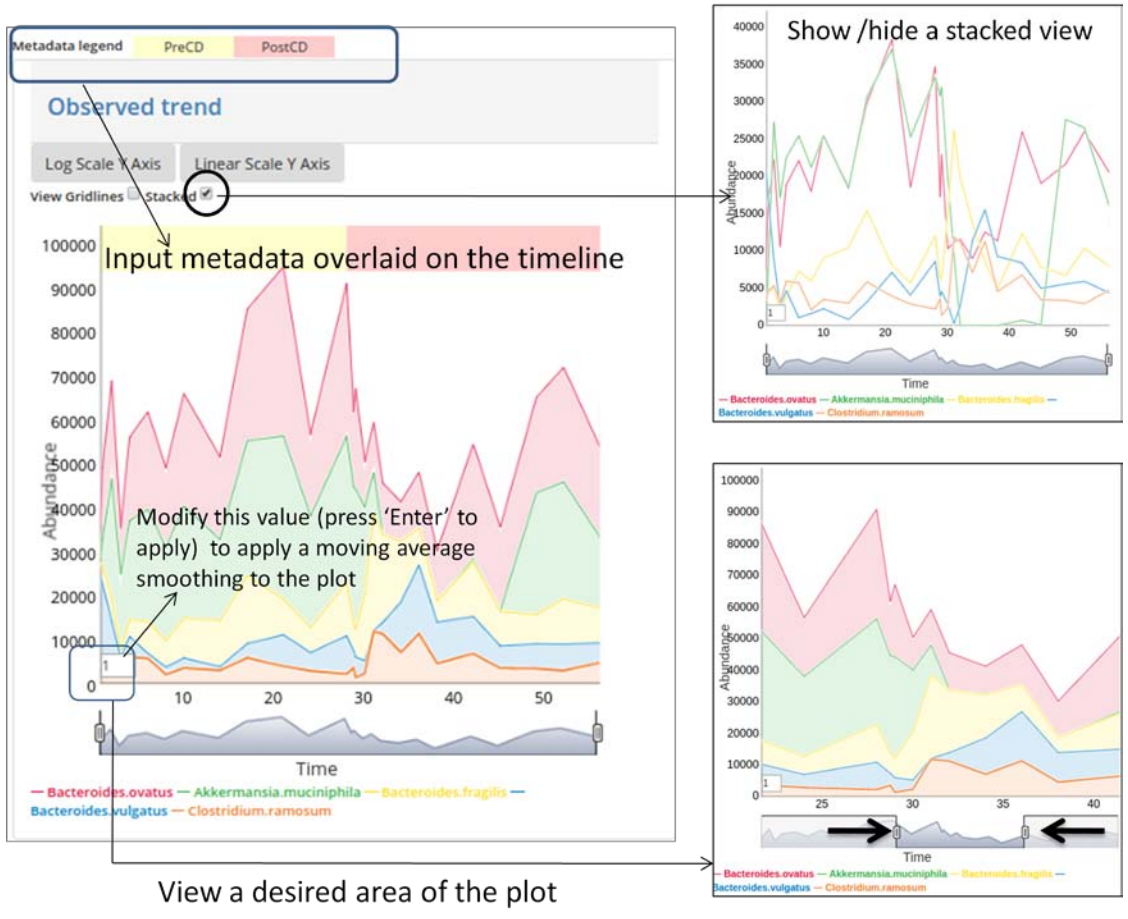
[Click here to show/hide correlation network](#)



Additionally, Web-gLV also generated an inter-taxa association network using the input taxa abundance matrix which can also be used as a guide to select a desired set of taxa for simulation. The network can be viewed by clicking on the link ‘Click here to show/hidden correlation network’. Clicking on a node selects it and its correlated nodes. Clicking on the Positive, Negative or All buttons will show only those specific types of associations. The thickness of an edge corresponds to the degree of its correlation. (N.B: The network is generated only with nodes having less than 30% zeroes and correlation  $\geq 0.5$  and  $\leq -0.5$ ). A quick selection can also be done by clicking the buttons labeled ‘select top 5’ or ‘select top 10’ abundant features. Alternately, the user can search a desired taxa name in the search-box or select it from the dropdown. The trend plots for added/removed taxa are automatically generated/hidden in the ‘Observed trend’ window. The trend plots for the ‘Predicted trend’ only appear after a successful simulation run.



Several user interactive operations like log transformation, stacking/un-stacking, viewing gridlines and selecting a desired window of the trend plot is possible. A moving average based smoothing can also be applied to all the time series plots by modifying the value in the left bottom corner box. Figures illustrating the various interactive operations are shown below:



## Step 2: Setting the simulation parameters

After selecting the desired set of taxa, the various parameters for setting up the simulation needs to be specified.

Parameter estimation settings

Specify start time point: Timepoint 1 (0.75)

Specify end time point: Timepoint 26 (56)

Enforce positive intrinsic growth and negative (or zero) self interaction constraints

Simulation settings

Time duration: 10 Interval: 0.1

Predicted Intrinsic Growth Rates

Initial values

Timepoint 1 (0.75)

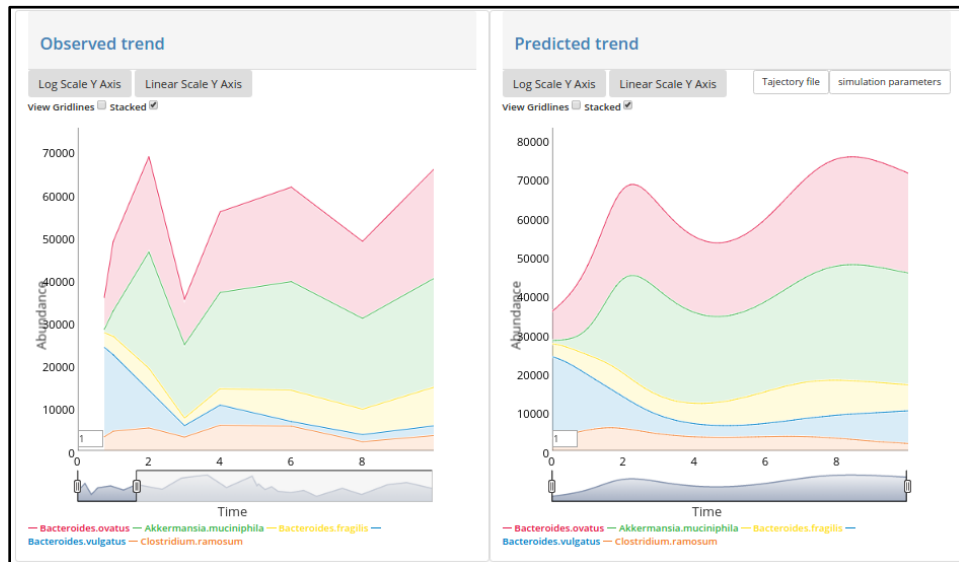
Selected initial seed values for starting the simulation (Click to edit values)

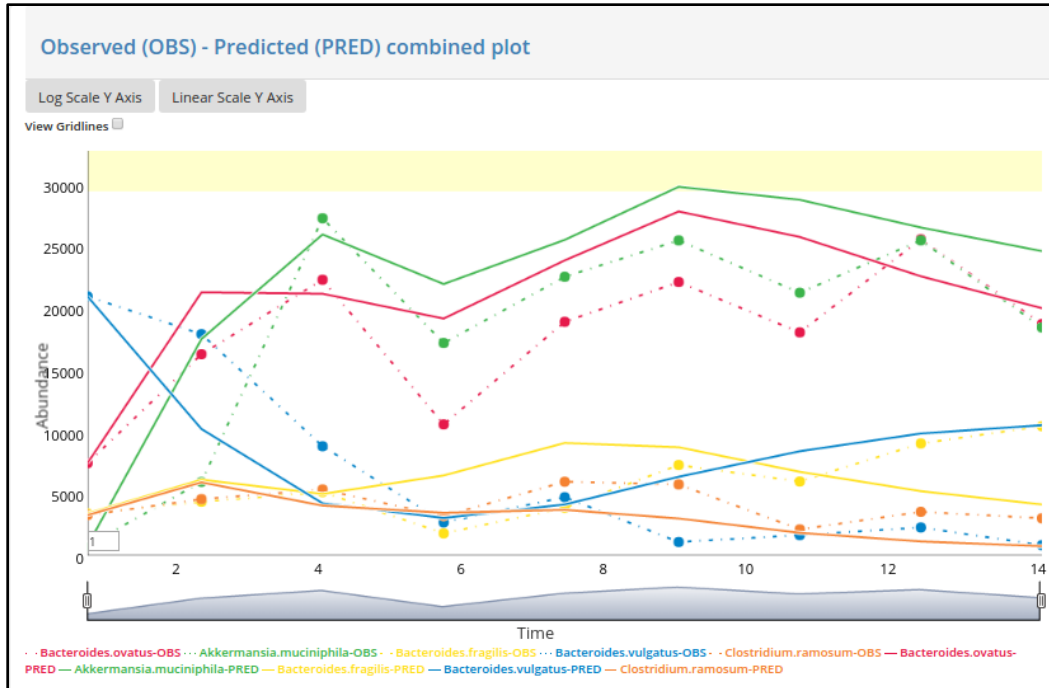
Name	Abundance
Bacteroides.ovatus	7454
Akkermansia.muciniphila	813
Bacteroides.fragilis	3391
Bacteroides.vulgatus	20904
Clostridium.ramosum	3259

The ‘start time point’ and ‘end time point’ corresponds to the subset of the input abundance matrix that would be used for parameter estimation. The number in the bracket beside the ‘Timepoint’ entry points to the actual sampling time-delta provided in the input dataset. The checkbox for constrained parameter estimation can be selected if a positive intrinsic growth rate and negative (or zero) self interaction constraint is desired. Selecting this option is suggested while estimating parameters from a microbiome dataset. Please note that these ‘Parameter estimation’ settings are not available when a simulation is started with a user supplied growth rate and interaction file. After setting the parameters, the simulation settings namely time duration, interval and the solver to be used can be specified from the available options. The initial seed values for starting the simulation can be selected as any one of the timepoints available in the input data. These values can also be modified to select a completely new starting point.

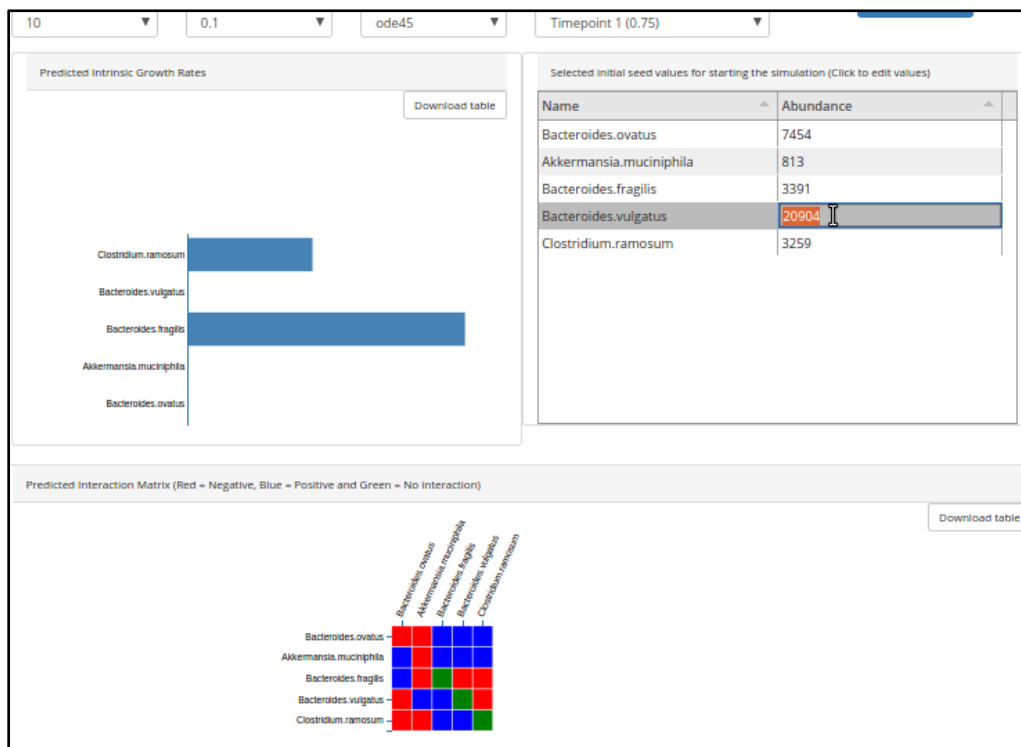
### **Step 3: Running the simulation**

After setting the parameters, The ‘Run simulation’ button can be clicked to perform a simulation. If the simulation is successful, the predicted trajectories for the selected taxa are displayed under the ‘Predicted trend’ window. In case of an unsuccessful simulation due to an incorrect parameter or solver limitation, an error message is displayed and no trajectories are generated. The timeseries plot in the ‘Observed trend’ window is automatically set to display the selected time range if the simulation range matches. This feature is helpful to compare the predicted trajectories from a modified starting point and compare it with the unmodified observed trends. The observed vs. predicted trend plot for a taxon is also generated as a mixed plot with the observed trends (OBS) shown in points connected by dotted lines and the predicted (PRED) as firm lines of same color. Mouse over on each line highlights the corresponding legend.





The predicted growth rates and interaction values used for the simulation are graphically displayed. A simulation can be re-run by altering some parameter/simulation settings as well as with a modified set of initial values.



## Step 4: Evaluating a simulation

The observed and predicted trajectories are compared using a Dynamic Time Warping algorithm (DTW). The distance is calculated between the scaled (between 0 and 1) time series belonging to the 'Observed' and 'Predicted' data which is presented as a table along with the trend plots in the web-gLV tool. Additionally, the 'all versus all' DTW distance is calculated for the 'Observed' and 'Predicted' data to generate the hierarchically clustered dendrograms. These dendrograms represent the temporal similarities between the selected microbial groups and hence a reflection of their community structure. A comparison of the dendrograms generated for the 'Observed' and 'Predicted' data can hence be used as a measure of the simulation prediction accuracy.

