

MLSTest Tutorial

Contents

| | |
|--|----|
| About MLSTest | 2 |
| Installing MLSTest | 2 |
| Loading Data..... | 3 |
| Main window..... | 4 |
| DATA Menu | 5 |
| View, modify and export your alignments | 6 |
| Alignment>viewer | 6 |
| Alignment> export | 6 |
| Distances | 7 |
| Managing Diploid sequences | 7 |
| Working with allelic profiles..... | 8 |
| Allelic Profiles>Make/View allelic profiles | 8 |
| Analysis for Allelic Profiles | 8 |
| BURST | 9 |
| BURST over all group definitions..... | 10 |
| Making Trees for sequences | 10 |
| Neighbor Joining Analysis (Concatenated NJ option) | 11 |
| The Tree Viewer | 12 |
| Scheme optimization..... | 13 |
| >Optimus Number of Loci | 13 |
| >Selecting a good combination of loci | 13 |
| Tests window..... | 14 |
| How to select a good combination: | 14 |
| Analyzing congruence among loci..... | 15 |
| Recommended steps for analyzing incongruence | 15 |

About MLSTest

MLSTest is a novel Windows based software for multilocus sequence data analysis. It is aimed to work with diploid sequences but it is useful for haploid sequences too. It constructs allelic profiles and makes basic clustering analysis with them, calculates discriminatory power with confidence intervals, typing efficiency and genotypic diversity. Additionally, includes tools for view, edit, convert and concatenate sequences. It uses several simple methods for tree construction (UPGMA, Neighbor Joining, Bio-Neighbor Joining, majority rule consensus) with the advantage of manage heterozygous sites. It calculates node support using bootstrap and approximated clade significance based on the Templeton test that is faster than bootstrap. Additionally, the software analyzes whether concatenation is suitable for the data set using different tests (bionj-incongruence length difference test, templeton test). It evaluates how the incongruence is distributed across the tree using a variation of the localized incongruence length difference test implementing a modified neighbor joining algorithm. MLSTest have tools to optimize and reduce MLST schemes, to select best loci for typing and analyzes whether extending the scheme will increase resolution. It calculates consensus networks and makes multidimensional scaling for visual clustering. Additionally, MLSTest looks for network structures in BURST graphs that which are evidence of recombination. Finally, it implements a user-friendly viewer for trees, BURST and multidimensional scaling graphs that allows several editing and exporting options.

Installing MLSTest

MLSTest requires a computer with Windows xp Service Pack 2 or better and Microsoft .net framework 3.5 or better.

You can download the installer from the next link:

v. 1.0.0.39:

<http://mlstest.codeplex.com/downloads/get/clickOnce/MLSTest1.0.application>

HOW TO INSTALL

- 1) Download the installer and click on the downloaded file. Follow the instructions. Installation is very easy. A warning window could be appearing saying that the publisher is unknown. Do not worry, this is because we can't sign the app. You can uninstall the app using the control panel.
- 2) MLSTest will look automatically for actualizations and inform to you if you want to actualize. We will use this method to fix bugs or to include some new features.

Loading Data

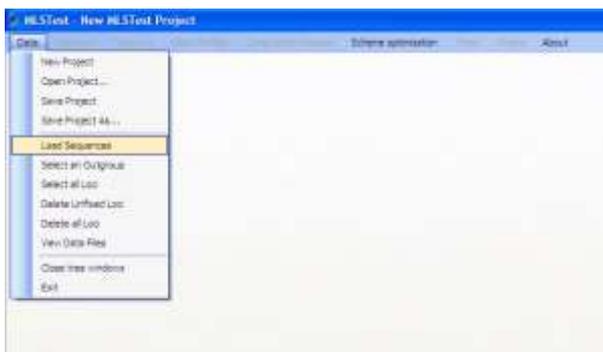
MLSTest just allows (at least for now) sequences in FASTA format:

Example of the format:

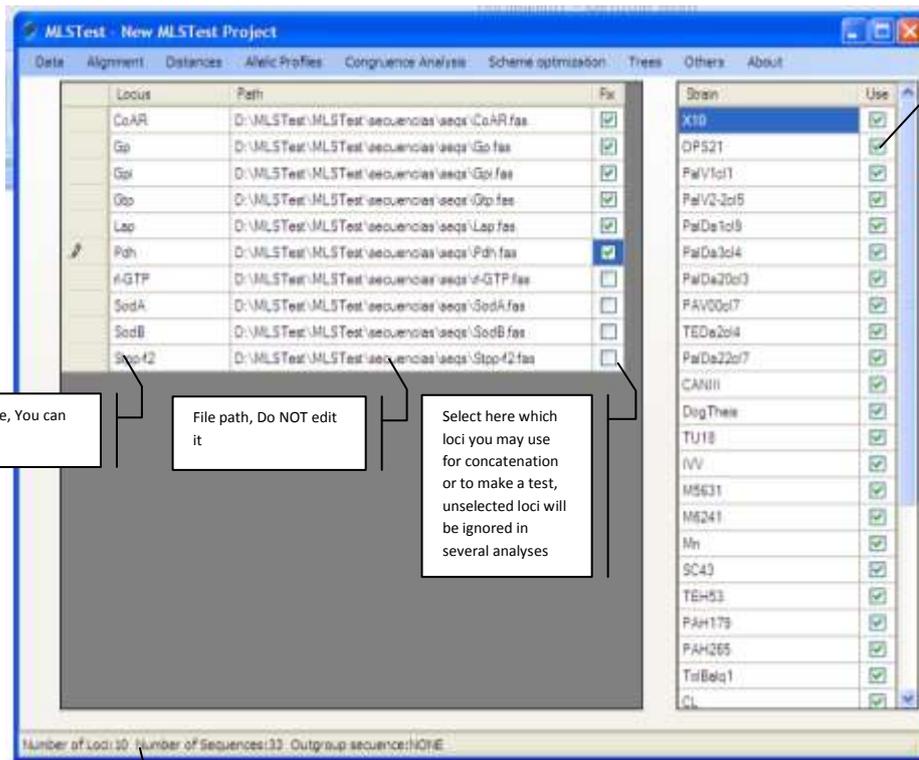
```
>sequence1
ATATTTTTTTTTT
>sequence2
ATTATRRTTTTTT
>sequence3
AGGGGGGGGG
```

IMPORTANT!!!! Sequences must be in the same order in each alignment loaded. MLSTest ignores names and consider that two sequences in the same position in different alignments correspond to the same organism.

First step: Load sequences



Main window



Locus Name, You can edit it

File path, Do NOT edit it

Select here which loci you may use for concatenation or to make a test, unselected loci will be ignored in several analyses

Here are showed strain names in your dataset. Select which strain are included in your analysis, unselected strains will be ignored

Main Information about your Data

DATA Menu

The image shows a screenshot of the 'MLSTest - New MLSTest Project' window. The 'Data' menu is open, displaying the following options: New Project, Open Project..., Save Project, Save Project As..., Load Sequences, Select an Outgroup, Unselect all Loci, Delete Unfixed Loci, Delete all Loci, View Data Files, Close tree windows, and Exit. Five callout boxes on the left provide descriptions for the first five menu items.

| Menu Item | Description |
|--------------------|---|
| New Project | It clears the current Project and opens a dialog to load alignments |
| Open Project... | Open a Saved project format .mls |
| Save Project | Save the current project. It is recommended. The paths of your alignments and your groups are saved |
| Save Project As... | Save the current project. It is recommended. The paths of your alignments and your groups are saved |
| Select an Outgroup | You may select an outgroup here. Your trees will be rooted with this sequence. |
| View Data Files | Show the main window when you are in other windows like the alignmet viewer |

View, modify and export your alignments

Alignment>viewer

Select here the alignment that you want to view

Select here if you want to view all sites or only the sites that have polymorphisms

If you edit a base here it will be changed in your files without confirmation. Be careful

Number of Loci:10 Number of Sequences:33 Outgroup sequence:NONE

Alignment> export

Click here to select which files export

Click here to select the format

Click here to save modified alignments

Distances

The Distances menu allows you to view and export distance matrix for alignments or allelic profiles. The file format of the matrix is compatible with phylip package for further analysis.

Managing Diploid sequences

Additionally, the user can select in “Distance>Handling Heterozygosities” how heterozygous sites are considered in the analysis. Three options are available:

1. Average States: the distance between two sites that are different is calculated as mean of distances among all the possible resolution of the heterozygosis

Example:

$$\text{distance C-Y} = \frac{\text{Distance C-C} + \text{Distance C-T}}{\text{Number comparisons}} = \frac{0 + 1}{2} = 0.5$$

Seq1: AATTY

Seq2: ACTTC observed distance=1 + 0.5= 1.5

2. SNP duplication and heterozygosis resolution (Tavanti et al, 2005): Constant sites are removed from the alignment. Then, polymorphic site are duplicated and heterozygosis are resolved.

Example:

Seq1: AATTY → AY → AAYY → AATC
Seq2: ACTTC → AC → CCCC → CCCC observed distance=3

3. Ignore in pairwise comparisons: Heterozigositities are not considered in pairwise comparisons

Seq1: AATTY → AATT
Seq2: ACTTC → ACTT observed distance=1.0

Only uncorrected distances are used by the software. However, considering the close relationships between strains in MLST datasets, probably this is not a problem because superimposed mutations should be rare. Additionally, this is a model-free approach, simpler and faster than corrected distances.

Working with allelic profiles

Allelic Profiles > Make/View allelic profiles

Allelic profiles are showed in a table. Additional information is included at the bottom of the table:

1. Number of Polymorphisms
2. Typing efficiency (TE). It is defined as the number of different genotypes (STs/DSTs) per polymorphic site.
3. Discriminatory Power: defined as the probability that two strains chosen at random from a population of unrelated strains will be distinguished by the typing method concerned (Hunter, 1990) and was calculated according to the following equation:

$$D = 1 - 1/N(N - 1) \cdot \sum_{j=1}^s x_j(x_j - 1)$$

where (s) is the number of types, (x_j) is the number of population members falling into the (j-th) type, and (N) is the size of the population.

The screenshot shows the MLST software interface with a table of allelic profiles. The table has columns for Name, CoAR, Op, Gp, Lsp, Pst, NGTP, SotA, SotB, SotC2, and DST. The rows list various loci such as TE1a2b4, TE1a2b7, etc. At the bottom of the table, there is a summary row for 'Number of Polymorph', 'Typing Eff', and 'Discriminatory Power'. Callouts point to specific parts of the interface: 'Locus Name' points to the 'Name' column, 'organism' points to the 'Name' column, 'Allele' points to the 'CoAR' column, 'ST' points to the 'DST' column, and 'Typing efficiency' points to the 'Typing Eff' row.

| Name | CoAR | Op | Gp | Lsp | Pst | NGTP | SotA | SotB | SotC2 | DST |
|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| TE1a2b4 | 5 | 8 | 4 | 3 | 6 | 5 | 5 | 1 | 1 | 7 |
| TE1a2b7 | 5 | 6 | 4 | 3 | 6 | 5 | 5 | 1 | 1 | 7 |
| TE1a2b8 | 8 | 7 | 5 | 4 | 7 | 6 | 6 | 3 | 2 | 8 |
| TE1a2b9 | 7 | 8 | 5 | 5 | 6 | 7 | 7 | 4 | 1 | 8 |
| TE1a2b10 | 8 | 9 | 7 | 6 | 8 | 8 | 6 | 5 | 3 | 10 |
| TE1a2b11 | 8 | 9 | 7 | 6 | 8 | 8 | 6 | 5 | 3 | 10 |
| TE1a2b12 | 9 | 10 | 8 | 7 | 10 | 9 | 8 | 6 | 4 | 11 |
| TE1a2b13 | 10 | 11 | 9 | 8 | 10 | 9 | 8 | 6 | 4 | 11 |
| TE1a2b14 | 11 | 12 | 10 | 9 | 11 | 10 | 9 | 7 | 5 | 12 |
| TE1a2b15 | 11 | 12 | 10 | 9 | 11 | 10 | 9 | 7 | 5 | 12 |
| TE1a2b16 | 11 | 12 | 10 | 9 | 11 | 10 | 9 | 7 | 5 | 12 |
| TE1a2b17 | 11 | 12 | 10 | 9 | 11 | 10 | 9 | 7 | 5 | 12 |
| TE1a2b18 | 11 | 12 | 10 | 9 | 11 | 10 | 9 | 7 | 5 | 12 |
| TE1a2b19 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b20 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b21 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b22 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b23 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b24 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b25 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b26 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b27 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b28 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b29 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b30 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b31 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b32 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b33 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b34 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b35 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b36 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b37 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b38 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b39 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b40 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b41 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b42 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b43 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b44 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b45 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b46 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b47 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b48 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b49 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b50 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b51 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b52 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b53 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b54 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b55 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b56 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b57 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b58 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b59 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b60 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b61 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b62 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b63 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b64 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b65 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b66 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b67 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b68 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b69 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b70 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b71 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b72 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b73 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b74 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b75 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b76 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b77 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b78 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b79 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b80 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b81 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b82 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b83 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b84 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b85 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b86 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b87 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b88 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b89 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b90 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b91 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b92 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b93 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b94 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b95 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b96 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b97 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b98 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b99 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| TE1a2b100 | 12 | 12 | 11 | 9 | 12 | 11 | 11 | 8 | 10 | 13 |
| Number of Polymorph | 37 | 26 | 28 | 48 | 30 | 30 | 38 | 23 | 18 | |
| Typing Eff | 0.381 | 0.52 | 0.428 | 0.228 | 0.453 | 0.375 | 0.245 | 0.256 | 0.476 | 0.222 |
| Discriminatory Power | 0.885 | 0.763 | 0.887 | 0.954 | 0.875 | 0.887 | 0.784 | 0.835 | 0.831 | 0.651 |

Analysis for Allelic Profiles

est Project

Distances Allelic Profiles Congruence Analysis Scheme optimization Trees Others About

CoAr Tcl prueba 2

1

4

5

5

6

5 5 6 4 4 8

2 8 8 2

7 4 6 7

4 9 4 3

8 4 4 3

5 6 6 8

Pdh To I plan A prueba 2

SodA Tcl prueba 2

ST

1 1 1

2 1 2

3 2 3

Basic BURST

BURST over all group definitions

Test for network structures

Copy to Clipboard

Export to file...

UPGMA-tree

NJ-tree

95% CI for Discriminatory Power

Select Just one Isolate per ST

Make/View Allelic Profiles

It is useful if you do not want to work with repeated STs (identical Strains)

It makes an UPGMA or Neighbor Joining tree for your allelic profiles

Calculates Confidence Interval for Discriminatory Power. It uses a delete-one jackknife procedure (Severiano et al., 2011). The CI is added next to

It makes a simple BURST analysis for your allelic profiles. You need to specify a group definition, usually the number of loci minus one

Calculates BURST over all group definitions in order to determine an optimal group definition. The percentage of singletons for each group definition and a dendrogram showing relationship among groups is showed

It analyzes whether network structures are present in BURST graphs suggesting recombination.

BURST

MSTest - New MSTest Project

Distances Allelic Profiles Congruence Analysis Scheme optimization Trees Others About

GROUP DEFINITION: Shared Alleles S/1E

Group 1: 1/1 of Strains 10 / Number of DST 2 / Predicted Four

DST FREQ STRAINS

16 2 F83, T6V23, T6F7, T6R6, T6V86, T6

18 1 CL

17 1 T6F6

Group 2: 1/1 of Strains 2 / Number of DST 2 / Predicted Four

ST FREQ STRAINS

10 1 R6C304

11 1 R6V507

Group 3: 1/1 of Strains 2 / Number of DST 2 / Predicted Four

DST FREQ STRAINS

10 1 T10

11 1 T11

Number of Loci: 10 / Number of Sequences: 33 / Outgroup base:

eBURST Graph

10 11

15

10 11

5 6

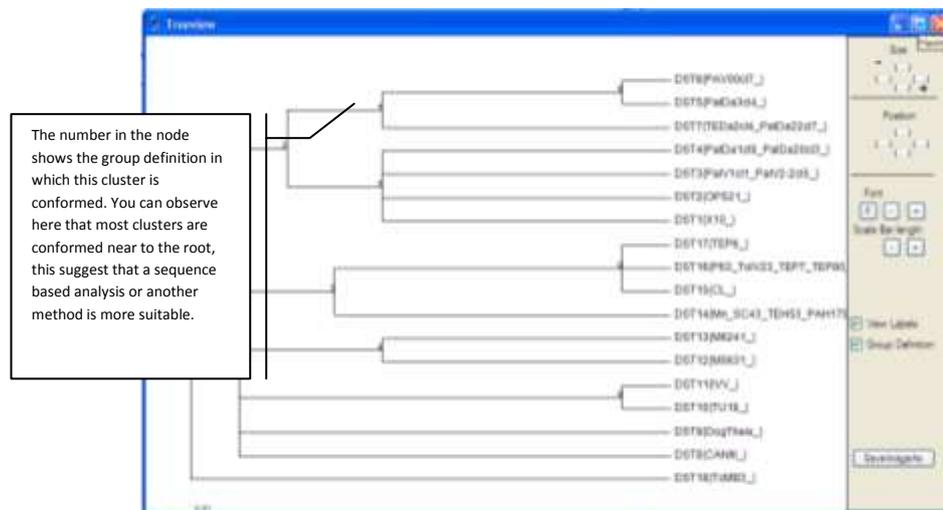
Option Bar for graph edition

Information about clusters conformed

Simple Graph with clusters, blue lined indicates STs connected by the group definition (number of shared alleles).

BURST over all group definitions

This analysis allow to see when groups are conformed using all possible group definitions. It is useful to select an optimal group definition or to decide if the eBURST analysis is useful for the data. For example, if the tree show the majority of groups conformed near to the root means that there are few DST/STs in common among strains. This suggests that an analysis not based on DST/STs is more suitable.

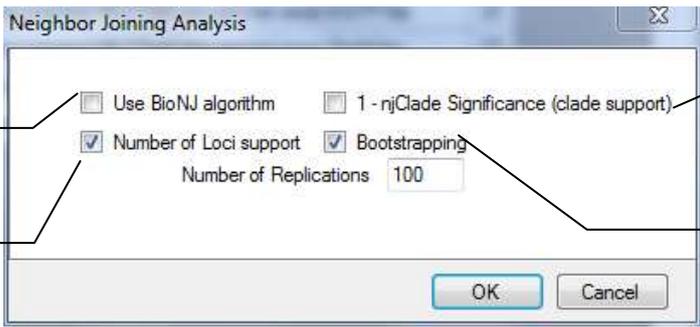


Making Trees for sequences

- 1) First, you need to specify how heterozygosities are treated in the analysis. Click on Distances>Handling heterozygosis and select one option. The default is Average States
- 2) Making a tree

The screenshot shows a software interface with a table of sequence files and a dropdown menu for tree construction options. The table lists various files with checkboxes in the 'Use' column. The dropdown menu is open, showing options like 'Concatenated-LQMA', 'NJ 50% NR Consensus Tree', 'IQ NJR Consensus Tree', and 'LQMA 50% Consensus Tree'. A callout box points to the 'Concatenated-LQMA' option, stating: "This option concatenates your selected loci and makes an UPGMA or Neighbor joining tree, you may specify if you want to make BioNJ tree or if you want calculate a support for each branch". Another callout box points to the 'NJ 50% NR Consensus Tree' option, stating: "Makes a NJ tree for every selected locus and then constructs a Majority rule consensus of the trees".

Neighbor Joining Analysis (Concatenated NJ option)



The image shows a software dialog box titled "Neighbor Joining Analysis". It contains several options and a text input field. Four callout boxes provide explanations for specific options:

- Use BioNJ algorithm:** Specify if you want to use a bioNJ algorithm instead of NJ.
- Number of Loci support:** Calculates the tree for every locus and then put in each branch of the concatenated tree the number of times that this branch appeared in individual trees.
- 1 - njClade Significance (clade support):** It calculates NJ based clade significance for selected groups or for all branches. Values are showed as 1-njClade significance (Clade Support).
- Bootstrapping:** Calculates bootstrap value for every branch, if you select this option a textbox is showed in order to specify the number of replications.

The dialog box also includes a "Number of Replications" text box with the value "100" and "OK" and "Cancel" buttons at the bottom.

The Tree Viewer

Tree Viewer

You can show several support measures at time separated by a bar

Right click on the branch to reroot, edit colors, edit font, select as group to test, calculate clade significance, show/hide, move, support.

Group name, it appears if you defined it in scheme optimization> groups to test

Changes the size of the tree

Moves the tree in the window or a selected element

Changes the font size and type

It moves all the Groups names and bars

Show/hide Group Names

Show/hide Taxa

Show/hide Support for all branches

If it is selected, when you click on a branch the tree is re-roted there

hide Support lower that the cutoff

This menu allows to color branches according their support value.

Save the tree in Newick Format or MLSTest Format

TEV67
TEV66
TEP80
TEP7
EPV20-1
EPP38
ToIV23
P63
CL
PAH179
TEH53
SC43
Mn
ToIBalq1
PAH265
TEP6

PaIV2-2cl5
PaIV1cl1
PaIDa20cl3
PaIDa1cl9
PaIDa22cl7
TEDa2cl4
PAV00cl7
PaIDa3cl4
X10
OPS21
DogTheis
CANIII
M6241
M5631
IVV
TU18
TcMB3

TcIV
TcII
TcIII

90/0
100/3
100/7
100/6
74/1
100/5
100/4
93/5
100/6
100/3
88/7
100/5
100/8

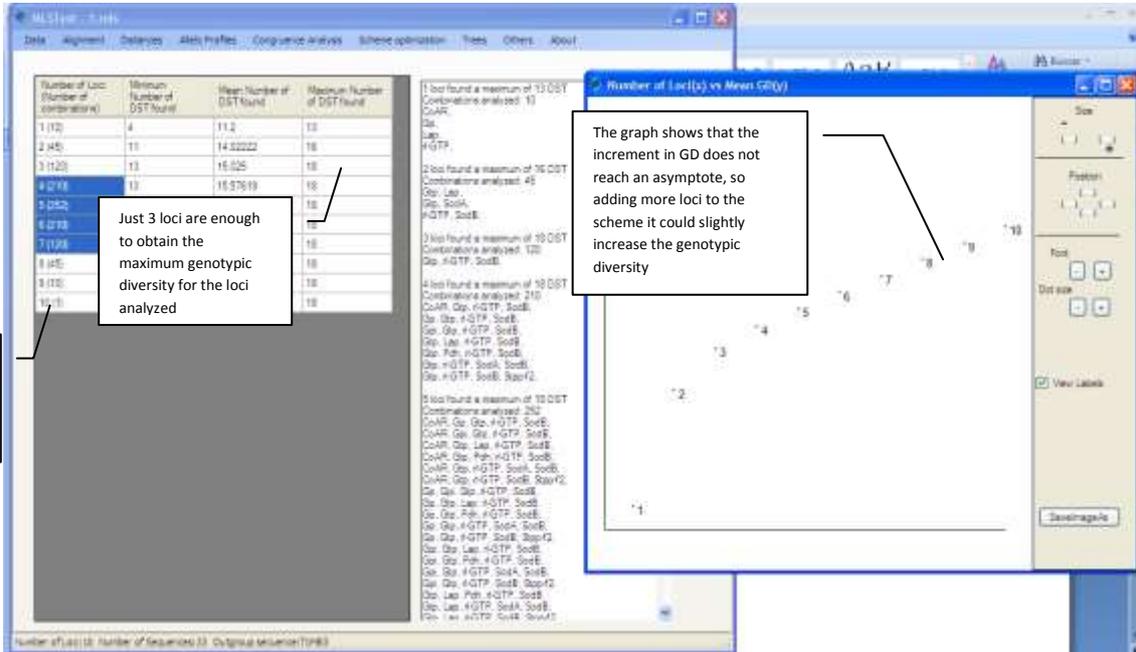
0.001

Size
Position
Font
Scale Bar length
Group Names
 View Node Names
 View Labels
 View Support
 Re-root
Branch colors
Support Cutoff
Bootstrap 70
 Show it
SaveImageAs
Export tree

Scheme optimization

>Optimus Number of Loci

This option is useful to reduce the number of Loci analyze and select a good number of loci in order to obtain a certain genotypic diversity. Additionally, it is useful to determine if the genotypic diversity will increase by adding more loci to the scheme.



>Selecting a good combination of loci

The screenshot shows the 'Scheme optimization' menu with several options. Callouts provide explanations for these options:

- Groups to Test:** Click to select groups to test if you previously know the phylogeny of your strains or based on the concatenated tree topology.
- Optimus Number of loci:** It makes the NJ trees for all possible combinations for a determined number of loci and evaluates and select the combination with better support and genotypic diversity.
- Test all combinations:** Add or delete one locus to a selected group of loci in order to test if there is a improvement or impairment in support, genotypic diversity or support.
- Stepwise addition:** Add or delete one locus to a selected group of loci in order to test if there is a improvement or impairment in support, genotypic diversity or support.
- Stepwise deletion:** Add or delete one locus to a selected group of loci in order to test if there is a improvement or impairment in support, genotypic diversity or support.
- Bootstrap Based Selection:** 1)Discard low supported loci, 2)calculate if a bootstrap for a branch in a combination is significantly better than the bootstrap of the same branch in another combination according to the number of replications used.
- Export tests...:** Save the tests for posterior view.
- Tests in the project:** Save the tests for posterior view.

Tests window

When you test all combinations of loci for a determined number of loci a window like this is showed. All possible combination of loci are showed in a table.

Click here to view tree if you selected "save trees" option

Genotypic diversity for the combination

Results about the different groups tested, "M" indicates that this group is monophyletic in the selected combination, "P" is not monophyletic.

Options that appear when you make right click over a combination. If you click "select these loci", the loci presented in the combination are selected in the main window.

Calculates different types of support for combination that are visible, non visible combinations will be deleted

Tools for select best combinations

Select the combinations that have the maximum number of tested groups as monophyletic

Select the combinations that have the maximum number of DST

Select the combinations that have the maximum number of tested groups with a support higher than a cutoff

Loci in the combination

How to select a good combination:

- 1) Define your selection criteria: i.e. maximize number of monophyletic groups and then Maximize observed diversity
- 2) Select the groups to test (scheme optimization menu>select group to test)
- 3) Run the analysis (scheme optimization menu>test all combinatios>test user determined groups. The tests window is showed
- 4) Apply the criteria clicking on "Maximize Monophyly", then click on "view" button. MLSTest applied this criterion and combinations that maximize the criteria, then click on maximize number of DST (the second criteria is applied over the first). So, combinations that maximize your criteria are showed

Analyzing congruence among loci

The screenshot shows the MLSTest software interface with the 'Congruence Analysis' menu open. The menu options are: BioNJ_Incongruence Length Diference, NJ_Localized Incongruence Length Diference, Templeton test, Topological incongruence, Tree Distance among loci, and Consensus Network. A list of paths is visible on the left, and several callout boxes provide detailed explanations for the menu items.

Analyze incongruence node by node of the topology of concatenated loci. A p value lower than 0.05 in a branch suggest that at least one locus is incongruent with the tested node. See Tomasini et al., for further detail.

It shows node by node the number of loci that are topologically incongruent with the concatenated loci tree

It shows an UPGMA tree of the Robinson-Foulds distance among loci trees. It is sometimes useful (especially when the number of loci is high in relation to the number of strains) to get an overview of topological incongruence among loci and which loci have similar trees.

It determines a p value for the BIONJ-incongruence length difference (ILD) for all selected loci. A p value lower than 0.05 indicates incongruence and the topology of the concatenated loci should be considered carefully

It tests if the selected locus/loci is/are incongruent to the concatenated loci tree using the Templeton test. A p value lower than 0.05 suggests the analyzed locus is incongruent with the topology of concatenated loci. The test can be made as an overall or a node by node analysis.

It makes a consensus network and save it. The consensus network allows viewing the topological incongruence in a network, alternatives branches are showed as squares or cubes. MLSTest cannot show the network graph in this version, but you can see it using Splitstree software.

About the Templeton test

Two tests are included based on the Templeton test. The first one, tests if the topology of the concatenated tree is not significantly different than the “optimum” topology for a selected partition based on the NJ method. It means that the data for this partition is incongruent with the topology of concatenated loci. This is made by evaluating both topologies site by site on the partition. The length of each topology for a site is calculated using the Pauplin formula (Pauplin, 2000). Then, the rank order of the differences is evaluated using the Wilcoxon sum-Rank test

The second test analyze node by node. It allows seeing how incongruence is distributed across the tree. Each branch in the topology of the concatenated is forced to exist in the data partition using the same algorithm described for NJ-LILD. Then, Templeton test is used to compare the optimal topology with the constrained topology. A p value lower than 0.05 indicates the partition is significantly incongruent with the evaluated node.

Recommended steps for analyzing incongruence

- 1) Look for topological incongruence (if there is topological incongruence go to the next step)
- 2) Make a BIONJ-ILD test, if the p-value is significant or near to the significance make a more detailed test as the NJ-LILD to find possible branches affected by the incongruence.
- 3) Test each locus to determine which is incongruent by using the Templeton test. You have two options here, to test the locus of interest against the concatenated topology or to test all the loci less the analyzed locus against the concatenated

topology. The first says if the data of the selected locus is incongruent with the combined topology. The second says if excluding certain locus, the remaining loci are poorly explained by the topology of the combined analysis. This means that this locus introduce significant topological modifications and sometimes it is powerful than the first approach. Sometimes, the incongruence is not detected in any single locus because it is distributed among several loci.