

MLVA-NET

Institut Pasteur MLVA database
www.pasteur.fr/mlva

The Institut Pasteur Multilocus VNTR Analysis (MLVA) database and web interface system (MLVA-NET) provides a common language on microbial strain typing based on MLVA data.

The purpose is to allow exchange of knowledge on the geographic and temporal distribution of strain types for epidemiology and evolution.

The main users are epidemiological surveillance networks and collaborative networks of microbiologists interested in population biology.

The MLVA-NET system was developed following the principle of the MLST database systems (www.mlst.net, pubmlst.org, www.pasteur.fr/mlst, web.mpiib-berlin.mpg.de/mlst/), which are widely used to access and search MLST databases.

MLVA data are often complementary to MLST data, as they provide discrimination among strains that are genetically homogeneous based on MLST. MLVA is now firmly established as one of the reference typing methods for outbreak investigation in a number of human pathogens such as *Mycobacterium tuberculosis*, *Bacillus anthracis* or *Salmonella enterica* serotype Typhimurium.

The MLVA-NET system provides a way to compare, search and download public MLVA data, or to access private datasets (for users with appropriate permissions). For each dataset, curators assign permission rights to registered users.

MLVA profiles are made public in order to provide the necessary common language, even though the corresponding isolate datasets can be defined as private in order to ensure confidentiality of epidemiological information or research projects.

One original feature of the system is that it is based on raw data (size of PCR products in bp). This harmonization feature enables data to be analyzed jointly, even when distinct profile coding methods are being used by distinct microbiologists.

At present (launch of version 0, October 2007), MLVA-NET only contains database query and profile comparison functions. In the near future, analysis tools (UPGMA, Minimum Spanning Tree, ...) will be integrated into the system.

Please contact the MLVA-NET administrators (S. Brisse and G. Guigon, pf8-bioinfo@pasteur.fr) for further information or to have a new organism or dataset added to the system.

The following sections include MLVA-NET definitions, the curator's manual and the user's manual.

MLVA-NET Definitions

Unregistered users

Unregistered users include any interested person that wants to consult and query profile information or isolate information in public datasets. Unregistered users are not identified by the system.

Registered users

Registered users are registered in the database and must identify themselves through login and password. Registered users can have access to private datasets as long as they have been given the rights to do so. Private datasets for which they have no rights are invisible to registered users.

Some limited information is registered for each user: their first and last name, affiliation and email address. New users that want to access private datasets must be declared in the system, in agreement with the curator(s) of the dataset.

Curators

Curators are registered users with special rights on one or more defined datasets and methods. They can add, update or delete entries in these datasets. They can also create, update or delete datasets and methods. Curators also define permission rights for each user on their datasets.

Organism

An organism is a group of isolates belonging to a common genus, species, subspecies and/or serotype (e.g., *Salmonella enterica* subsp. *enterica* serotype Typhimurium). Different datasets (private or public) can be attached to the same organism, e.g. by distinct groups of collaborating laboratories.

Isolates

An isolate is an entry of the database with attached source information and raw data. Raw data are fragment sizes (in base pairs) obtained for each studied marker. Each isolate belongs to a single dataset, and its information (name, country, year, source...) will only be available to authorized users. However, the corresponding profile information will be immediately and automatically made public for harmonization purposes.

Dataset

A dataset is a set of isolates (e.g., studied by one laboratory or one collaborative network) with specific access rights. A dataset can be either public or private. Public datasets are accessible to everyone, including unregistered users. Private datasets are password-protected and only accessible (and visible) to defined users with appropriate rights.

There are 3 possibilities of access rights for users on datasets:

- No rights: the user cannot access the dataset, nor even know about its existence.
- Consultation only: the user can consult the isolates information and their associated profiles.
- Consultation and curation: the curator(s) can add more isolates to the dataset, and update or delete isolates.

Each dataset is handled by one or more authorized curators, who can add new isolates, update information or delete the dataset. They can also give rights to other users on the

dataset. A dataset is linked to a reference method, which defines the set of markers used to study the isolates.

Method

A method is a way to study isolates' raw data. It consists of a set of markers (VNTR regions chosen for the study), each with a set of intervals (bins) specifying the allelic number corresponding to a fragment size range. The markers order is also specified in the method. Thus, a method is the translation of raw data into MLVA profiles.

Each method is attached to a specific organism. However, an organism can be studied by several methods, which can differ by their marker set, intervals, order, or a combination thereof. This feature provides the unique possibility to analyze the same raw data in a user-specific way.

Profiles

A profile is the combination of allele numbers for each isolate. An isolate will have one profile for each method that is linked to its dataset. The profile is characterized by a repeat type (RT) number. This number is unique for a given method; the same profile code in two distinct methods have no special affiliation (e.g., RT 1 in method A and RT 1 in method B need not correspond to the same isolate).

Allelic numbers may correspond to the deduced number of repeats, or may correspond to arbitrary numbers defined chronologically (as is the case for MLST alleles). If one or more raw data value is missing, the corresponding allelic number will be '999' and the corresponding profile will be considered as incomplete. RT numbers of incomplete profiles will be preceded by the letter 'd' and numbered consecutively, without affiliation to complete profiles (e.g., RT 1 and RT d1 have no particular affiliation). Query functions can be restricted to complete profiles, to incomplete profiles, or instead can combine all profiles.

If a raw data value does not fall into a predefined bin (size interval), the corresponding allelic code will be '1000' and the profile will be considered as incomplete (with '1000' being considered distinct from '999'). It is up to the curator to determine how to handle these special cases.

For harmonization purposes, it is important that isolates sharing the same profile for a given method end up with the same profile code. Hence, profiles are public data, available for every user. But their attached private isolates remain inaccessible to unauthorized users (see rights on datasets).

MLVA-NET Curator's manual

To insert data into the MLVA-NET database, three objects are necessary: the organism (e.g., 'S. enterica ser. Typhimurium'), the method (e.g., 'NIPH method'), and a dataset (e.g., 'NIPH isolates'). Once these are defined, isolates can be entered.

To access to curator page, click on the button [<Access to curator page>](#) on the home page (www.pasteur.fr/mlva). You will have to identify yourself with your login and password. Only curators are allowed to access this interface.

The curator index page provides (i) The entire list of datasets for which you have curator rights, and (ii) The available actions on these datasets.

New organism

To have a new organism added to the database, please contact the administrator of MLVA-NET (pf8-bioinfo@pasteur.fr).

New method

If it does not already exist, the first thing to do is to add the method you want to use:

[<Insert new method>](#)

- You now have to select the organism, and add information about the new method: name, description, the number of markers, and sender (the id of the user who is proposing this new method).

Insert new method

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Organism: !	<input type="text" value="salmonella enterica enterica typhimurium"/>
Name: !	<input type="text" value="Pasteur Method"/>
Description: !	<input type="text" value="The STTR markers are ordered according to their number (STTR3, STTR5, STTR6, STTR9, STTR10). Bins were slightly modified compared to NIPH method."/>
Number of markers: !	<input type="text" value="5"/>
Sender Id: !	<input type="text" value="1"/> Click for a list of sender Ids

[<Submit>](#)

- On the second page, you must select the markers used. If they already exist in the database, just select their name in the menu, next to the marker's number. In this case it is not useful to fill in the form for this marker (it will not be used). If your marker is new, enter its name and information.

Now please select the markers used by this method, in the correct order.
 If the markers are new, enter informations about them - required fields are marked with an exclamation mark (!).

Marker 1:

Fill in the fields only if the marker is new:

Name: ! (ex: STTR5)

Length: ! (ex: 6)

Sequence: ! (ex: ACCACG)

Offset: ! (ex: 181)

Gene name:

Primer 1: !

Primer 2: !

Marker 2:

Fill in the fields only if the marker is new:

Name: ! (ex: STTR5)

Length: ! (ex: 6)

Sequence: ! (ex: ACCACG)

Offset: ! (ex: 181)

Gene name:

Primer 1: !

Primer 2: !

<Submit>

- On the next page, you must add the intervals for each marker: the minimal fragment size value (bin_min), the maximal one (bin_max) and the corresponding allele number. You should copy your data from your tab-delimited spreadsheet and paste them in the correct marker's text field.

Add Intervals
 Now paste in your tab-delimited text from your spreadsheet for each marker. Click here for [example data](#).
 Do not include the header line and ensure that all fields are in the correct order (bin_min, bin_max, allele number).

Try to insert wide intervals to avoid values falling between 2 intervals.

Marker 1: STTR3

279	309	5
312	342	6
345	375	7
378	408	8
411	441	9
444	474	10
477	507	11
510	540	12
543	573	13

Marker 2: STTR5

322	326	24
328	332	25
334	338	26
340	344	27
346	350	28
352	356	29
358	362	30
364	368	31
370	374	32

<Submit>

If it passes the automatic check, the new method will be inserted into the database. Subsequently, all isolates with available raw data for all markers included in the new method will be automatically coded according to the new method. Consequently, these isolates will be immediately available for study using this method.

New dataset

<Add new dataset>

After selecting the organism, you will be requested to provide information about the dataset: name, description, status (private or public), and a reference method to be used by default. This default method will define the available markers for the dataset.

Optionally, you may link some publications to the dataset. To do so, please submit their name and URL (tab-delimited) so that they can be accessed directly through the website.

First select the isolate's organism:

salmonella enterica enterica typhimurium If you don't find the organism you want, please [insert a new one](#)

Then fill in this form - required fields are marked with an exclamation mark (!).

Name: ! Pasteur Isolates

Description: ! Isolates studied at the Institut Pasteur

Status: ! private

Reference Method: Select the method and its associated markers used to study this dataset.
(All methods using the same markers will be linked to the dataset).

NIPH method: STTR9 STTR5 STTR6 STTR10 STTR3

Pasteur method: STTR3 STTR5 STTR6 STTR9 STTR10

If you don't find the method you want, please [insert a new one](#) first.

Enter the dataset's reference's name and URL: Paste in your tab-delimited text from your spreadsheet.
First column = reference's name / Second column = reference's URL

Schuffenecker, I. et al. <http://dx.doi.org/10.1371/journal.pmed.0030263>

<Submit>

- You will next be asked to define permissions. If your dataset is private, you will need to define permissions for users and curators. If your dataset is public, only curator rights can be defined (users automatically have consultation rights).

Now please give permissions to users and curators:

- Brisse Sylvain
- Cheval Justine
- Guigon Ghislaine

New isolates

To add new isolates in one of your datasets, two ways are available. You can add a single isolate with the « add new isolate » function, or add several at the same time with the « batch insert » function.

<Add new isolate> To insert your isolates one by one, just fill in the form with isolate information and raw data for each marker. Required fields are marked with an exclamation mark (!).

Add new isolate

First select the isolate's organism:

salmonella enterica enterica typhimurium If you don't find the organism you want, please [insert a new one](#)

Then select the isolate's dataset:

Pasteur isolates If you don't find the dataset you want, please [insert a new one](#)

Then fill in this form - required fields are marked with an exclamation mark (!).

Strain name: !

Other name 1:

Other name 2:

Serotype:

Phage type:

pfge:

atb:

Source:

Country:

Year:

Outbreak:

Comment:

Source Lab:

Sender: ! [Click for a list of sender ids](#)

And enter the raw data for each marker (0 for null value):

STTR3: !

STTR5: !

STTR6: !

STTR9: !

STTR10: !

<Batch insert> To insert isolates in batch, copy/paste from a spreadsheet (e.g. Excel). A tab-delimited header for your spreadsheet is available. When pasting your values, do not include the header line.

Batch insert

First select the isolate's organism:
 salmonella enterica enterica typhimurium

Then select the isolate's dataset:
 Pasteur isolates

Please paste in your tab-delimited text from your spreadsheet. Do not include the header line and ensure that all fields are in the correct order.
 (strain, other_name1, other_name2, serotype, phage type, pfge, atb, source, country, year, outbreak, comment, source lab, sender, rawdata for each marker).
 Isolate's id, dataset, curator and dates will be automatically added.

Markers order: STTR3, STTR5, STTR6, STTR9, STTR10

												<input type="button" value="Reset"/>
isolate 1												
BBPE_Pasteur	Cluster	D,Patient	Typhimurium									
1	519.00	265.00	352.00	171.00	France	104	X1	A,C,S,Sp,Su,Te	2000		Patient13	
isolate 2												
BBPE_Pasteur	Cluster	D,Patient	Typhimurium									
1	490.30	0	352.60	188.80	France	104	X1	A,C,S,Sp,Su,Te	1998		Patient14	
isolate 3												
BBPE_Pasteur	Cluster	D,Patient	Typhimurium									
1	517.00	265.00	358.50	170.60	France	104	X1	A,C,S,Sp,Su,Te	2001		Patient16	
isolate 4												
BBPE_Pasteur	Cluster	D,Patient	Typhimurium									
1	0	234.80	0	0	France	104	X1	A,C,S,Sp,Su,Te	1999		Patient22	
isolate 5												
BBPE_Pasteur	Cluster	D,Patient	Typhimurium									
1	250.00	253.00	346.70	170.60	France	104	X1	A,C,S,Sp,Su,Te	2002		Patient25	

Isolates with correct values will be inserted into the database. Their profiles obtained with the different methods will be automatically created (if they are new) or deduced (if they already exist).

The allele numbers obtained with the reference method will be printed in a table. If an isolate with the same name already exists in the database, a warning message will be issued and the isolate will not be inserted again. If in addition, the raw data are different for the same isolate name, an error message will be printed. The curator should then check which of the two entries has correct raw data. In the same way, if a value is submitted with a wrong format (for example, non-numerical raw size data), an error message is printed and the isolate is not inserted. Note that isolates with more than two null data (this criterion may evolve later on) will not be accepted in the database, as illustrated in the Figure below.

New isolate(s) added:

4 isolates have been added to database. In the following table you can check their allele numbers obtained with the dataset's reference method (Pasteur method). (Note that if you have '1000' values, your fragment sizes do not correspond to any interval. The reference method should be updated with new intervals.)

id	Strain	STTR3	STTR5	STTR6	STTR9	STTR10	RT
201	isolate 1	12	14	15	4	23	1
202	isolate 2	11	999	15	6	10	d31
203	isolate 3	12	14	16	4	24	22
204	isolate 5	1000	12	14	4	21	d32

Format error:

Line	Error message	strain	other name1	other name2	serotype	phage type	pfge	atb	source	country	year	outbreak	comment	source lab	sender	STTR3	STTR5	STTR6	STTR9	STTR10
4	more than 2 null raw data entered	isolate 4			Typhimurium	104	X1	A,C,S,Sp,Su,Te	Cluster D,Patient	France	1999		Patient22	BBPE_Pasteur	1	0	234.80	0	0	362.90

If you stop the insertion process before the end, no data will be inserted in database.

Query – Update – Delete

Curators can also query their data, and then update or delete them.

Note: the query menu is only available if the curator has already submitted some data.

To **query data**, fill in some specific fields in the query page (method, dataset or isolate query page) or, to retrieve all records, leave all fields blank and click on <Submit>. A table with all matching data will be printed. It allows you to check your data, and then **delete** or **update** some of them.

Dataset query/update

Searches will match any field where an entry has been made. Leave a field blank to return all matches for that field. Leave all fields blank to retrieve all records.

Organism:	salmonella enterica enterica typhimurium	Reset
Id:	<input type="text"/>	
Name:	<input type="text"/>	
Private:	<input type="text"/>	
<input type="button" value="Submit"/>		

Delete	Update	Id	name	private
Delete	Update	1	Pasteur isolates	1

All three types of data (method, dataset and isolate) can be queried or deleted, but only information about datasets and users can be updated. You will be allowed to modify the users rights on your dataset, and change the dataset name, description and reference method. You can also add new publications and delete some others.

Name:	<input type="text" value="Pasteur isolates"/>
Status:	<input type="text" value="private"/>
Description:	<input type="text" value="Isolates studied at the Institut Pasteur"/>
Reference Method:	<input type="radio"/> NIPH method: STTR9 STTR5 STTR6 STTR10 STTR3 <input checked="" type="radio"/> Pasteur method: STTR3 STTR5 STTR6 STTR9 STTR10
References:	<p>Enter the dataset's reference's name and URL: Paste in your tab-delimited text from your spreadsheet. First column = reference's name / Second column = reference's URL</p> <input type="text"/>
Select the references you want to delete:	<input type="checkbox"/> Schuffenecker, I. et al. : http://dx.doi.org/10.1371/journal.pmed.0030263
<input type="button" value="Update"/>	

If you realize you made a mistake in the insertion of an isolate, you will have to delete it and to insert it again with the correct data. Likewise, to change or add intervals in your method, first delete the method and then insert it again with the new data. These features are for traceability purposes, as new IDs will be given to new isolates or methods.

As a curator, you should carefully check that all data are correct. In addition, please check that your method does not imply too many values that fall outside the predefined bins. In this case, it would be better to redefine the method by adding more intervals or enlarging them

MLVA-NET User's manual

Users can consult either profile data or isolate data.

In both cases, please proceed by first selecting the organism studied, and then select either <profiles> or <isolates>.

Institut Pasteur MLVA database

Please select the organism you want to work on:

Salmonella enterica ser. Typhimurium ▾

and select the information you want to consult :

For **profile** queries, now choose the coding method to be used, and then <submit>. Information about methods is available on the same page.

Please select the coding method:

- NIPH Method:** The method initially published by Lindstedt and al., 2005. Marker order is STTR9, STTR5, STTR6, STTR10, STTR3.
- Pasteur Method:** The STTR markers are ordered according to their number (STTR3, STTR5, STTR6, STTR9, STTR10). Bins were slightly modified compared to NIPH method.
- mynewmethod:** test

The query functions are now accessible. The number of available profiles is given at the end of the page. Note that most queries can be performed on either complete profiles, or incomplete profiles, or both.

Welcome to the *Salmonella enterica* subsp. *enterica* serotype Typhimurium profiles database

Method used : NIPH Method. [Click to change method.](#)

This server provides the following tools for querying the profiles database:

Allelic profile / repeat type queries

- [Single allelic query](#) - find all RTs with a given allele.
- [Profile query](#) - find all RTs that are identical or similar to a given profile.
- [Batch query](#) - assign RTs to multiple profiles copied from a spreadsheet.
- [Repeat type query](#) - find RTs that have similar allelic profiles to your query RT.
- [Batch Profile coding](#) - from fragment sizes to allele numbers.
- [Search database](#) - advanced queries.
- [Browse database](#) - browse and export all records.

Downloads

- [Download allelic profiles](#) - download all known allelic profiles in tab-delimited text format.

Database tools

- [Set options](#)

Database statistics

- Number of profiles: 119

For **isolates** queries, please first log in if you are a registered user with access to private dataset(s).

To access private datasets, please log in:

Login

Password

Now please select the method and the available dataset(s) you want to study. Note that datasets with different marker sets cannot be studied together. Information about methods is available on the same page.

Choose one of the coding methods and select one or more corresponding datasets:

NIPH Method: The method initially published by Lindstedt and al., 2005. Marker order is STTR9, STTR5, STTR6, STTR10, STTR3.

Pasteur isolates (public): Isolates studied by the Pasteur Institute

PasteurTestPublic (public): To test the curator interface

Pasteur Method: The STTR markers are ordered according to their number (STTR3, STTR5, STTR6, STTR9, STTR10). Bins were slightly modified compared to NIPH method.

Pasteur isolates (public): Isolates studied by the Pasteur Institute

PasteurTestPublic (public): To test the curator interface

On the next page, the different possible queries are accessible.

Welcome to the *Salmonella enterica* subsp. *enterica* serotype Typhimurium isolates database

Method used : NIPH Method. [Click to change method.](#)

Selected dataset : Pasteur isolates.

This server provides the following tools for querying the isolates database:

Allelic profile / repeat type queries

- [Repeat type query](#) - find RTs that have similar allelic profiles to your query RT.
- [Profile query](#) - find all RTs that are identical or similar to a given profile.
- [Browse database](#) - browse and export all records.
- [Search database](#) - advanced queries.

Database tools

- [Set options](#)

Database statistics

- Number of isolates: 170
- Last updated: 2007-10-15

Diversity indices

A number of diversity indices, including Simpson's index, can be calculated on selected sets of isolates for each marker individually and for allelic profiles.

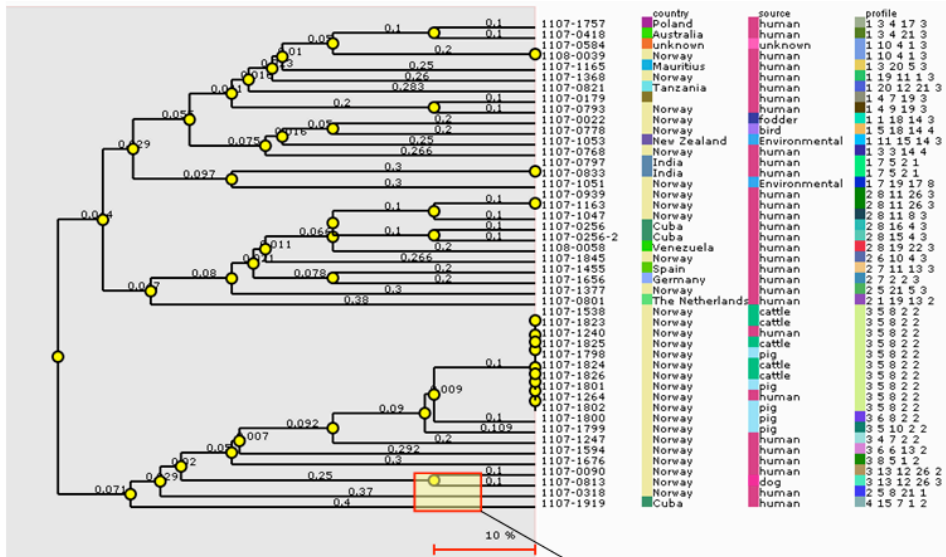
	Simpson	Shannon (H')	Shannon (E)	Nel	Number of Isolates
STTR9	68.89%	1.1690	0.8432	0.67	46
STTR5	86.76%	2.2001	0.8578	0.85	46
STTR6	90.63%	2.5079	0.8852	0.89	46
STTR10	85.12%	2.1429	0.8624	0.83	46
STTR3	62.51%	1.1175	0.6944	0.61	46
RT	95.36%	3.2377	0.9181	0.93	46

Evolutionary models

Distances between profiles can be calculated using several evolutionary models such as the saltational model or by applying (on user-defined loci) the stepwise mutation model (SMM), which considers alleles with similar repeat sizes as being more closely related. For the SMM model, the user must define two parameters: the minimal coefficient (MC), which corresponds to the weight of the most similar alleles (one repeat difference) relative to the weight between distant alleles. The second parameter is the saturation level (S). It corresponds to the repeat number difference beyond which the alleles are considered equally distant. For example, with $MC = 0$ and $S = 3$, alleles that differ by only one repeat will not contribute to profile distance, while alleles that differ by 2 repeats contribute with weight 0.5 to the overall profile distance.

Dendrograms

Interactive UPGMA and Neighbor-joining trees can be displayed with attached isolate information for user-defined fields. There are three possible layouts: phenogram, circular or radial. Choosing an isolate as outgroup for re-rooting is possible. The tree can be exported in the Newick format. Branch collapsing is obtained by clicking on a node. Display of distance and strain information is optional. In the radial and circular displays, strain information can be obtained by placing the mouse over the color circles. Allelic profiles are displayed when placing the mouse over the strain name.



- id
- rt
- strain
- serotype
- phage_type
- source
- country
- year
- sourcelab
- date_stamp
- profile
- Reset
- Select all
- Show/Hide Distances

Strains selected from the tree will appear in the text field below

[Tree picture : PNG format](#)

532	115	1107-1919			human	Cuba	2007	NIPH Oslo	2007-12-12
4	15	7	1	2					
228	88	1107-0318			human	Norway	2007	NIPH Oslo	2007-11-14
2	5	8	21	1					
195	84	1107-0090			human	Norway	2007	NIPH Oslo	2007-11-14
3	13	12	26	2					
388	95	1107-0813			DT104	dog	Norway	2007	NIPH Oslo
3	13	12	26	3					

Future developments

Development of additional analysis and visualization tools is currently ongoing. Please consult the web site for updated information.