User Guide for MLVAbank 6.0

FOR MICROBES GENOTYPING

Version 1.3.1 april 2016

This document is the sole property of the *Institut de Génétique et Microbiologie*, UMR8621, University Paris-Saclay. Any representation or reproduction, in whole or in part, made without the consent of the *Institut de Génétique et Microbiologie* is forbidden. 2007-2013, Orsay.

TABLE OF CONTENT

1. Introduction and Basics 3
2. Getting started
2.1. Creating an account 3
2.1.1. Difference between guests and
users 3
2.1.2. Create an account 4
2.1.3. Login 4
2.2. Editing account information 4
2.3. Creating a group 4
2.4. Editing a group 5
2.5. Join a group 5
2.6. Forgotten password5
3. Managing Databases 6

3.1. Creating a database	.6
3.2. Browsing Databases	.7
3.3. Modifying and deleting	.8
3.4. Exporting and Importing via CSV	.8
3.4.1. Exporting	.8
3.4.2. Importing	.9
3.5. Accessing geographic view 1	0
3.6. Querying a database1	0
3.7. Handling Panels and Genotypnumbers 1	
3.7.1. Panels 1	2
3.7.2. Genotype numbers 1	3

1. Introduction and Basics

The aim of this site is to facilitate microbes (including pathogenic bacteria) genotyping essentially for epidemiological surveillance purposes. Typically, **MicrobesGenotyping** will be used to set-up a genotyping assay appropriate for specific needs. The data will be derived *in silico* from whole genome sequence data. Low cost and low resolution typing may be useful as part of routine quality-checking of strain identity, or as a first-line exclusion assay in an outbreak investigation.

Typing is sequence-based. Any data which fits in a table, as derived for instance from assays such as polymorphic tandem repeat typing (MLVA for Multiple Loci VNTR analysis) or sequence typing (MLST for Multiple Loci Sequence Typing) can be managed.

More specific information on individual databases and associated assays can be found on dedicated support websites (links to articles, protocols, primer information).

WHAT IS A DATABASE IN MICROBESGENOTYPING?

A database is an interactive representation of typing data. **MicrobesGenotyping** can support any numerical data. A database can be queried or viewed. It can be shared with other registered users or even made public.

WHAT IS A QUERY?

A query is a comparison of a genotype (a string of allele values) to the genotypes present in a database. A database is selected, then a set of markers representing an assay is chosen, before submitting data from one isolate. The result lists the closest strains present in the database. The distance between two genotypes is the number of different alleles (categorical or Hamming distance). A Newick format tree can be exported for drawing in dedicated software, as well as the matrix distance or genotyping data for the closest strains.

WHAT IS A VIEW?

A View displays all strains present in a database. Fields can be hidden. Lines can be sorted according to field content. The database creator or administrator may have made accessible extra general information on the database on an independent web site, the URL will be indicated.

WHAT IS A CSV FILE?

A CSV file (comma-separated values) stores tabular data in a plain-text form. Each record contains a number of fields, separated by a comma ("," most used format or ";" french-style format).

2. GETTING STARTED

2.1. Creating an account

2.1.1. Difference between guests and users

Guests (no account required) can query and view public databases. A logged-in user can create, import, query, view and manage databases, including sharing them with other users, or making them public.

2.1.2. Create an account

To create an account, click on the "Sign Up" button in the home page menu bar, and follow the instructions. Registration is free. If the account creation is successful, you will be redirected to the "Edit Profile" page and you will receive an email.

2.1.3. Login

Account owners can click on "Log In" and fill the login form with their credentials.

2.2. Editing account information

Click on username in the "users" menu (Figure 1).



Figure 1: accessing the profile page

The left-end form manages login credentials (email, password) and the right-end is optional profile information. Forms need to be saved separately.

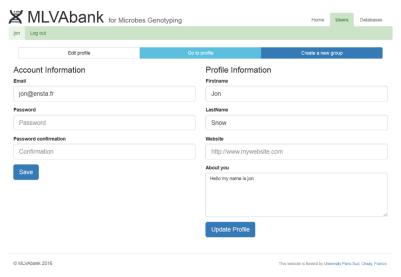


Figure 2: editing account and profile information

2.3. Creating a group

To create a group, click on username in the users menu (Figure 1) and then on the "create a new group" tab. Provide name, description and choose the members of this group (Figure 3; the creator is automatically included but can be removed). Alternatively, a group can be created while importing a database.

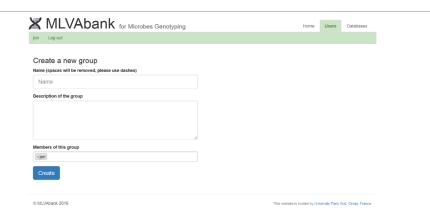


Figure 3: creating a new group

2.4. Editing a group

Members of a group are allowed to edit that group. Access the editing page by clicking on "my databases" in the database menu and then on "edit" next to the group name.

Members of this group

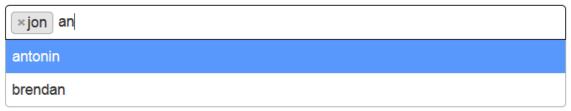


Figure 4: editing group members

Name and description can be changed, users can be added and removed (Figure 4).

2.5. Join a group

New members can be added by a current member.

2.6. Forgotten password

Click on "Forgotten password" on the login page and fill the form with the email associated with the account. The password will be reset and an email with the new password will be sent.

3. MANAGING DATABASES

3.1. Creating a database

Log-in, select "Databases > Create New Database". Choose a CSV file to upload (choose ";" or "," depending on CSV file format).

It is not possible to use different delimiters (for instance tab and ",") in the same CSV file.

Tip: export data from a public database to obtain an example CSV file.

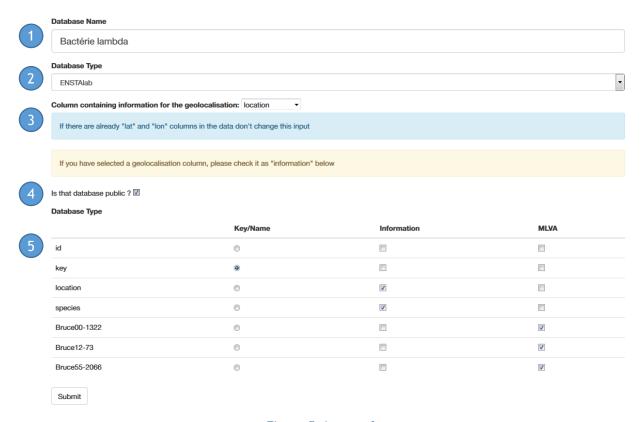


Figure 5: import form

Indicate (Figure 5):

- 1. Database name
- 2. the group can be a pre-existing group or a new group. Select "personal" if the database is not shared within a group.
- 3. the geolocalisation information field (cf Accessing geographic view)
- 4. the status (public or not) of the database
- 5. the type of each field to be imported from the CSV file. One field must be defined as the key and contain a unique identifier for each strain. The other fields can be either data (that must be an integer) or information (such as localisation, host, contact, etc.). Fields without type assignation will not be imported.

Press "Submit" to create the database. It may take a while if there are a lot of data.

Note: field names should not contain special characters (such as <, >, /, ...) or the corresponding column will not be maskable.

3.2. Browsing Databases

Public databases can be viewed, queried and exported with no need to log-in. Login is needed to create and access private databases.

The databases listed can be sorted by clicking on the header of the column.

In "My databases", toggle (show/hide) database lists, to access the group of interest (Figure 6).

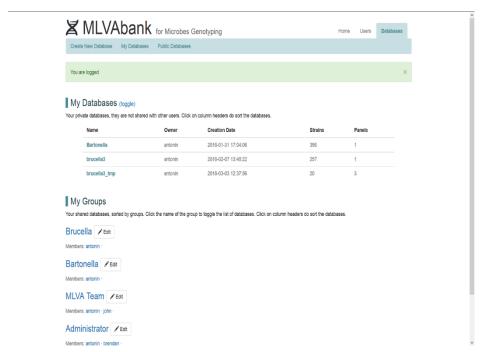


Figure :6 database listing page of a logged user

Each database has its specific header including database name, owner, description and link for more support if available (Figure 7).

See the database's menu to edit, query, view, etc.

Show/hide columns for clearer navigation.

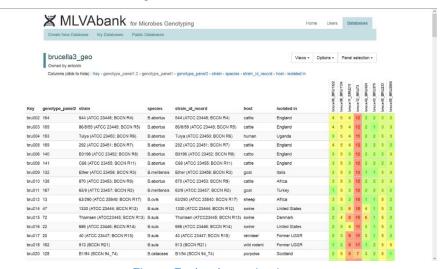


Figure 7: database viewing

3.3. Modifying and deleting

A database can be edited by its owner via the "Options > Edit" menu (Figures 7 and 8).



Figure 8: database editing page

Click "Edit" to save changes.

A database can be deleted by its owner via the "Options > Delete" menu. Confirmation will be required.

Once deleted, all metadata (information fields) and genotyping data are lost.

3.4. Exporting and Importing via CSV

MicrobesGenotyping does not provide internal data editing tools. Instead, CSV files are used to modify data, including adding or deleting strains. When importing, new fields (informations or typing data) can be added.

Users able to view a database can export it. Only users belonging to the database-associated group are allowed to import data in the database.

3.4.1. Exporting

To export, select "Options > Export" in the database menu.



Figure 9: database export page

The database export page (Figure 9) allows to select information fields to be exported. A data fields selection can be applied by selecting a specific panel (use option "None" to export only a selection of the information fields).

The CSV delimiter can be selected according to local settings.

If checked, "Advanced Export" option will add extra lines at the beginning of the file, defining field types and panels. These lines are recognised by **MicrobesGenotyping** and will facilitate subsequent re-import.

For more information on genotype read paragraph <u>Handling Panel and Genotype numbers</u>. Press "Export" when done.

3.4.2. Importing

To import data in a pre-existing database, select "Options > Import" in the database menu.



Figure 10: database import page

Choose a CSV file to upload(1) and the delimiter ";" or "," depending on the file's format (2) on the database import page (Figure 10).

Choose modifications to apply during the import from the CSV file to the database (3). You can import partial files, the only column needed is the "key" column that will allow the website to identify the strains.

- a. "Add new strains" to add strains with a key that is not already in the database, use that option to update a database or merge it with another one that share most fields with it.
- b. "Update old strains" to update existing strains (same key). Fields (columns) that are not present in the CSV file will remain unchanged.
- c. "Add new columns" to add columns that are in the CSV file but not yet in the online database. They may be categorized as "information" or "mlva" later, as done during database creation. This option only creates empty columns, the (a) and/or (b) options also need to be checked to fill them.
- d. "Add panels" to import panels. Use a file resulting from an "Advanced Export" as an example to define panels in the CSV file.

The last two options in Figure 10 are accessible to the database owner only

e. "Import genotypes number" cf Handling Panel and Genotype numbers

f. "Delete strains that are not in the CSV". All deleted strains are lost and cannot be recovered, so use this option very carefully.

3.5. Accessing geographic view

If the database includes geographic location, click on views -> Strains geolocalisation to access the map.

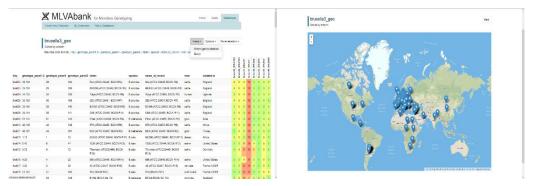


Figure 11: geolocalisation of strains

Note: About Geolocalisation information

Geolocalisation needs to be defined during database creation.

- a. If there are two fields (columns) named lat and lon, they will be used automatically
- b. In the absence of **lat** and **lon**, select the field containing the geographic information, latitude and longitude will be computed during the creation. **Also assign the field to the "information" type.**

If there are two columns "lat" and "lon" but information is missing for some strains, option 2 can be used in combination: when "lat" and "lon" information is available, it will be used. If not, the field defined as geographic information field will be used. In a subsequent export, the "lat" and "lon" fields will be filled with the deduced values.

3.6. Querying a database

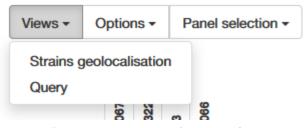


Figure 12: accessing the query form

To query a database, go to its page (e.g. "Databases > Public Databases") then "Views > Query" in the database's menu (Figure 12).

The query takes into account the panel chosen when viewing the database.

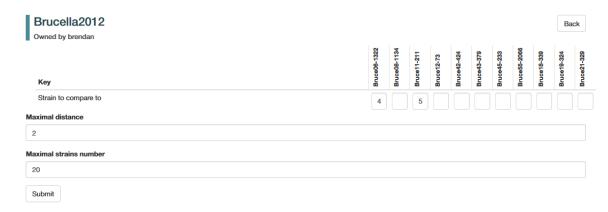


Figure 13: submitting a query

Set the genotype to be compared to genotypes in the database by pasting the whole genotype in the leftmost box (Figure 13). The values will be automatically parsed. Alternatively enter values in each box. Choose the "Maximal distance" between the query and the strains (Hamming distance) (e.g. 2) and the "Maximal number of strains" to be listed in the results page (e.g. 20). Press Submit. Expect a longer processing time if high "Maximal distance" and "Maximal number of strains" output values were selected (Figure 14).

wned by brenda	an	hard dealers		to bloom one	A.P No.							Export •	Back	ıck
Distance to queried strain	e - Key - strain	- host - isolated	host	ies-biovar - year - pu	species- biovar	year	publication		group	MLST9	lon	lat	Bruce06-1322	
	Reference strain												4	
I	2006LeFl	BCCN#94-19	cattle	Limoges, France	B. abortus	1994	Le Flche2006	Gilles Vergnaud	B.abortus		1.261105	45.833619	4	
I	2006LeFI	BCCN#94-18	cattle	Limoges, France	B. abortus	1994	Le Flche2006	Gilles Vergnaud	B.abortus		1.261105	45.833619	4	
I	2006LeFI	BCCN#95-34	cattle	Sicilia, Italia	B. abortus	1995	Le Flche2006	Gilles Vergnaud	B.abortus		14.0153557	37.599993	8 4	
I	2006LeFl	BCCN#95-31	cattle	Sicilia, Italia	B. abortus 4	1995	Le Flche2006	Gilles Vergnaud	B.abortus		14.0153557	37.599993	8 4	
I	2006LeFl	BCCN#99-98	cattle	Mongolia	B. abortus 7	1999	Le Flche2006	Gilles Vergnaud	B.abortus		103.846656	46.862496	4	
I	2006LeFl	BCCN#93-15	cattle	Pamplona, Spain	B. abortus	1993	Le Flche2006	Gilles Vergnaud	B.abortus		-1.6482653	42.811663	1 4	

Figure 14: query output page

The query result is shown in a way similar to database viewing. Strains can be sorted by clicking on column headers and show/hide informations columns. The geographic origin of matching strains can be displayed if the database contains geolocalisation information.

The "export" menu allows to draw a dendrogram.

The genotyping data of matching strains can be exported as a CSV as previously shown for the whole database by selecting "Export > CSV".

The distance matrix (categorical distance) can be retrieved by clicking on "Export > Distance Matrix". To export it to the MEGA format, click on the button "Export to MEGA format" (Figure 15).

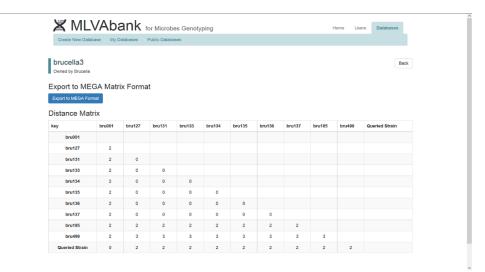


Figure 15: distance matrix

A Newick Tree can be obtained by clicking on "Export > Newick Tree". The Newick tree code will be at the bottom. Several links (buttons at the bottom) also allow to display the tree.

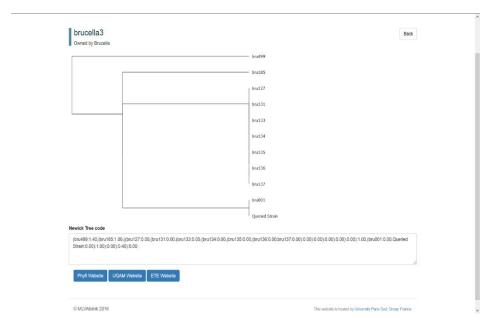


Figure 16: UPGMA tree

3.7. Handling Panels and Genotype numbers

3.7.1. Panels

Panels are made to view and query databases more easily. Most importantly, panels can correspond to different types of assays, for instance MLVA, MLST, canSNPs...

Select a panel in the "Panel selection" tab from the database's menu. Members of the database's group can access the "Handle Panels" menu to add or edit panels.

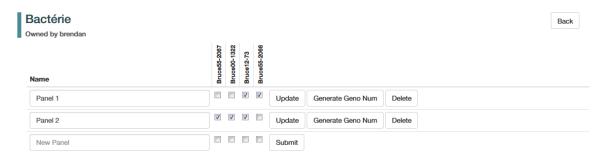


Figure 17: panel editing page

Check or uncheck the marker in the corresponding line and press "Update" to validate. A panel name can be changed, given that each panel has a different name. Changes must be validated before editing an other panel.

To delete a panel, click on the "Delete" button of the corresponding line.

To create a panel, edit the last line and press "Submit" to validate.

3.7.2. Genotype numbers

To facilitate the quoting of a genotype, genotype numbers may be assigned. It is then common to all strains sharing that particular genotype.

To generate genotype numbers for a particular panel, go to the "Handle Panels" page and click the "Generate Geno Num" button. It will create genotype numbers for every strains, those numbers are marked "temp" as they must be verified by an user.

Genotype numbers can be modified by exporting with the "Advanced Export" option checked, editing, and importing back (cf Exporting and Importing via CSV). It is not necessary to indicate the new genotype number for all strains, as the import tool will spread genotypes to any strain sharing the same genotype. The "Add new strains" and/or "Update old strains" need to be checked in order to load the genotype numbers for the new and/or the old strains.

When exporting a database as a CSV file, genotype numbers that are not yet verified can be exported. After re-importing, the genotype numbers will be verified for all users. Be careful on that option.