Microbes Genotyping User Guide - version 1.40

User Guide – version 1.4.0

This document is the sole property of the Institut de Biologie Intégrative de la Cellule (I2BC), UMR9198, University Paris-Saclay. Any representation or reproduction, in whole or in part, made without the consent of the *I2BC* is forbidden. 2007-2018, Orsay.

Table of content

1 Introduction and basics			4
		ing started	4
	2.1	Creating an account	4
	2.1.1	Difference between guests and users	4
	2.1.2	2 Create an account	4
	2.1.3	B Login	5
	2.2	Editing account information	5
	2.3	Creating a group	5
	2.4	Editing a group	6
	2.5	Join a group	6
	2.6	Forgotten password	6
3	Man	aging databases	7
	3.1	Creating databases	7
	3.2	Browsing databases	8
	3.3	Modifying and deleting	9
	3.4	Exporting and importing via CSV	9
	3.4.1	Exporting	9
	3.4.2	2 Importing	10
	3.5	Accessing geographic view	10
	3.6	Querying a database	11
	3.7	Handling Panels and Genotype numbers	13
	3.7.1	Panels	13
	3.7.2	2 Genotype numbers	14

1 Introduction and basics

The aim of this site is to facilitate microbes (including pathogenic bacteria) genotyping essentially for epidemiological surveillance purposes. **MicrobesGenotyping** makes data from a genotyping assay easily accessible. The data may 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 prior to whole genome sequencing.

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) or Single Nucleotide Polymorphism (SNP) 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?

A database is an interactive presentation of typing data. **MicrobesGenotyping** can support any numerical data. A database can be queried, viewed or exported. It can be shared with other registered users or 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 results list the closest genotypes 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.

Whats 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, view and export public databases. Inadditional ogged-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.

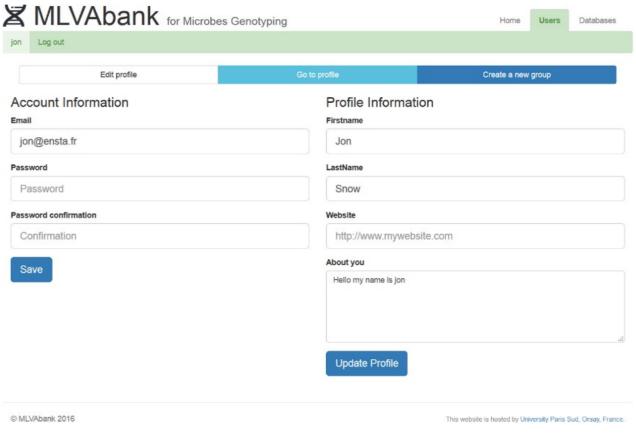


Figure 2: Editing account and profile information

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.

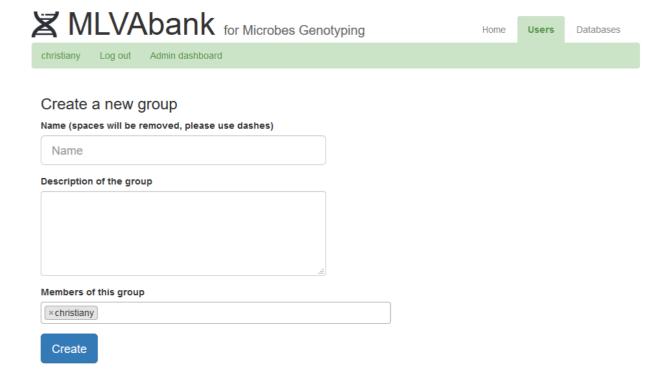


Figure 3: Creating a new group

2.4 Editing a group

Members of a group are allowed to edit the group. Access the editing page by clicking on "my databases" in the database menu and then on "edit" next to the group name. Name and description can be changed, users can be added and removed (Figure 4).

Members of this group

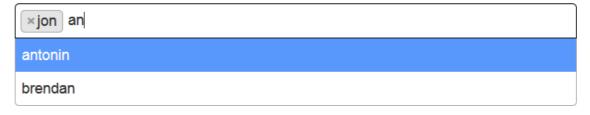


Figure 4: Editing group members

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 databases

Log-in, select "Databases > Create New Database". Choose a CSV file to upload. It is not possible to use different delimiters (for instance tab and ",") in the same CSV file. Press "Submit" to create the database. It may take a while if there are a lot of data. Fields can be used to create hyperlinks (contact e-mail address, URL link to publication or sequence file, etc.).

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

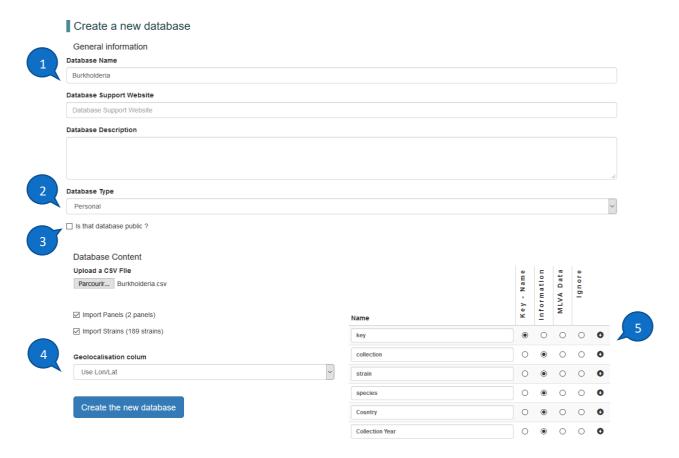


Figure 5: Import form

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 status (public or not) of the database
- 4. the geolocalisation information field (cf Accessing geographic view)
- 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 metadata (such as geographic origin, host, contact, etc.) or typing data. Fields without type assignation will not be imported.

Note: field names should not contain special characters (such as <, >, /, ...). Link to another site, publication, or email address are defined using parentheses: contact_name(example@email.com). Avoid special character(s) such as "é", "è", "i" etc in "contact_name".

3.2 Browsing databases

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

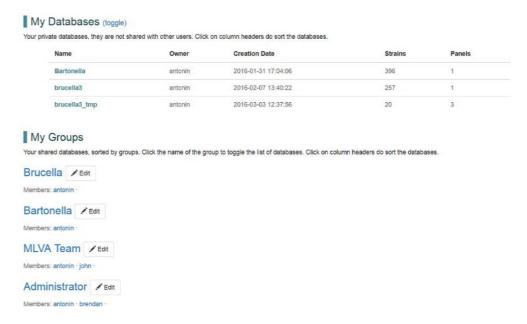


Figure 6: Database listing page of a logged user

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.

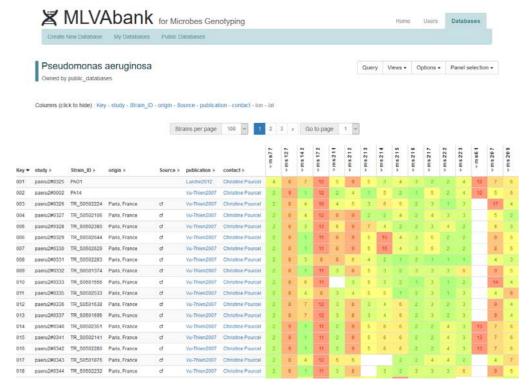


Figure 7: Database viewing

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 (metadata 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 "Views > 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. Press "Export" when done.

For more information on genotype read paragraph Handling Panels and Genotype numbers Importing.

3.4.2 Importing

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

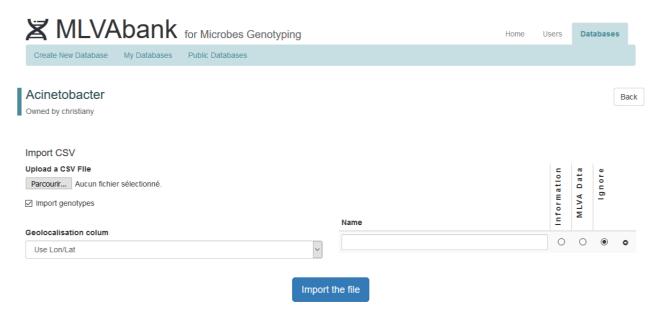


Figure 10: Database import page

Choose a CSV file to upload on the database import page (Figure 10). Partial files can be imported, the only column needed is the "key" column that will allow the website to identify the strains.

The geolocalisation column is selected as during the creation step (Creating a database). The "Import genotype number" option in Figure 10 is accessible to the database owner only (cf Handling Panel and Genotype numbers)

3.5 Accessing geographic view

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

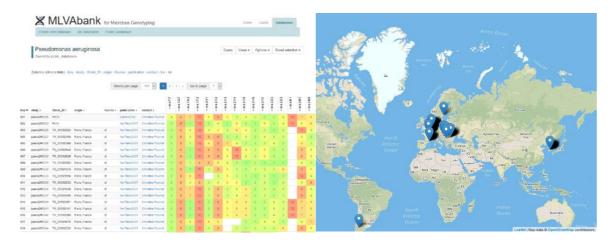


Figure 11: Geolocalisation of strains

Geolocalisation needs to be defined during database creation. It can be done in two ways:

- If there are two fields (columns) named lat and lon, they will be used automatically
- 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

To query a database, go to its page (e.g. "Databases > Public Databases") then "Query" in the database's menu (Figure 12). The query takes into account the panel chosen when viewing the database.

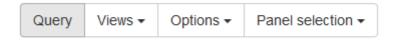


Figure 12: Accessing the query form

The different alleles can be typed individually or the whole genotype can be pasted in the leftmost box. The values will be automatically parsed. 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). The "only one strain per genotype" option allows to show one strain per genotype (the first found). Press Submit.



Figure 13: Submitting a query

Expect a longer processing time if high "Maximal distance" and "Maximal number of strains" output values were selected.

The query result is shown in a way similar to database viewing (Figure 14). Strains can be sorted by clicking on arrows next to column headers and show/hide metadata columns. The geographic origin of matching strains can be displayed if the database contains geolocalisation information. The genotyping data of matching strains can be exported as a CSV as previously shown for the whole database by selecting "Views > Export as CSV".

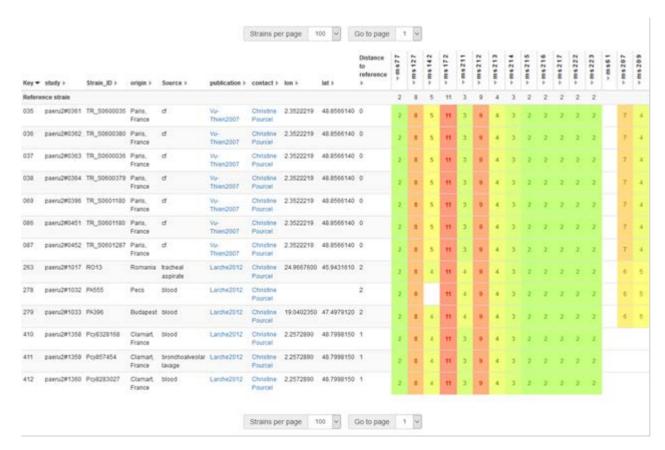


Figure 14: Query output page

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

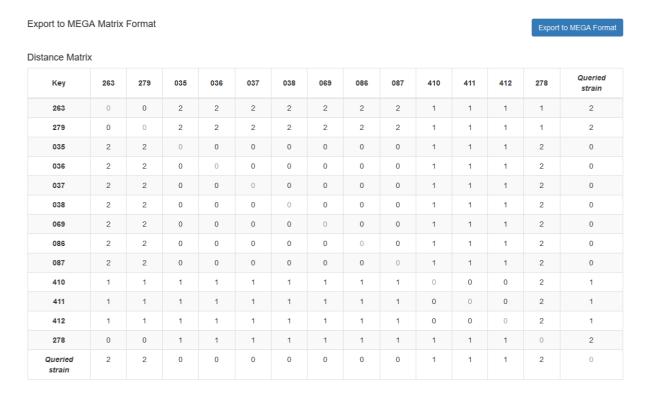


Figure 15: Distance matrix

A Newick Tree can be obtained by clicking on "Views > Newick Tree". A representation can be selected from the drop-down menu: Radial, Diagonal, Circular, Rectangular and Hierarchical. Labels can be modified as well from the second drop-down menu.

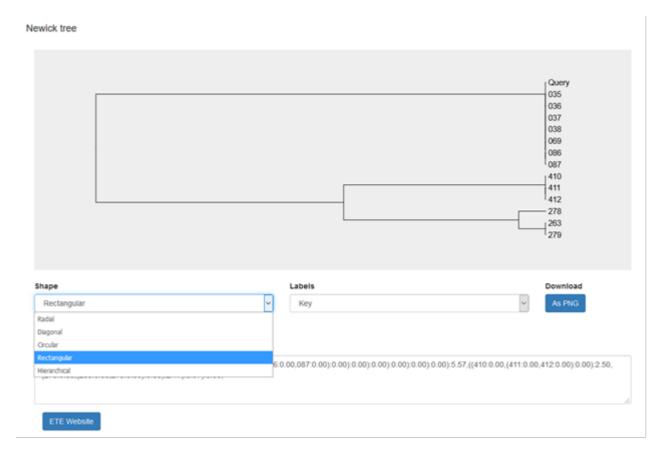


Figure 16: UPGMA tree

A tree can be downloaded as jpeg file with the download button or as Newick format tree code at the bottom. The "ETE website" button is a link to ETE toolkit, a phylogenetic tree viewer.

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, etc. 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.

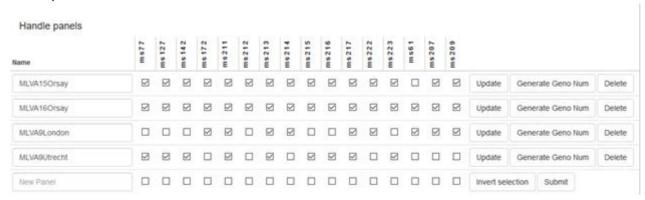


Figure 17: Panel editing page

To modify an existing panel, 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, you can inverse the selection by clicking on the "Invert selection" button.

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.

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.

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.