

ForeST Manager Manual

Version 1.0

US FISH AND WILDLIFE FORENSICS LABORATORY

August 22, 2020

Contents

1 ForeST Manager Architecture	3
1.1 Installation on Windows 10	3
1.1.1 Instructions once you have installed everything in the previous section but you want to run the application again.	10
1.1.2 Database	11
2 OMFF-ForeST Manager interface	13
2.1 Database manager	14
2.2 ForeST consult	17
2.3 Consensus spectra	20
3 ForeST web page update for Arbor Harbor	25

APP context

This program was written as a web APPI for management of data from DART obtained from wood samples.

Author and help

This program was designed and created by Sandra Martinez Jarquin, any question, issue or bug you find you can contact me for help. E-mail: jarquinsandra@gmail.com. Original code can be pulled directly from my git repository for developing purposes.

1

ForeST Manager Architecture

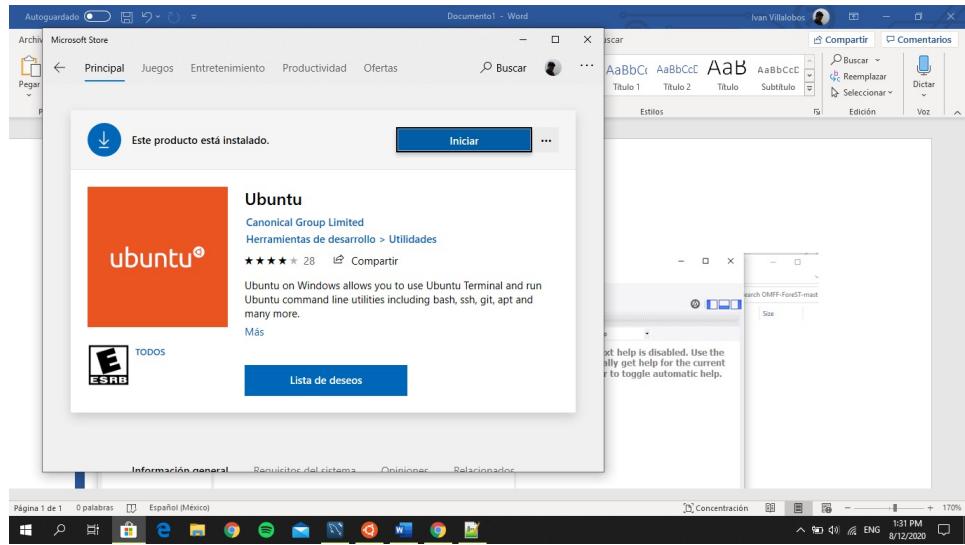
ForeST Manager was build as a Flask application, using Python 3.7.1, as a web application, but it can run in a local server with the database configuration. Packages used to build the original application and their versions are included in the file requirements.txt, however updates maybe necessary when using a different python version. The app was build using a blueprint approach, with three blueprints db manager, search and landing. the config folder was used for different stages of development, and can be used to deploy the application to a web server or a docker container if required. The app has been uploaded to a github repository <https://github.com/jarquinsandra/OMFF-ForeST>, and this manual refers specifically to local installation on a windows server.

1.1 Installation on Windows 10

For installing the application on windows 10, an Windows Linux System (WLS) should be used, if you have already installed a WLS Ubuntu system you can skip to the point seven, if you have installed everything on this section once on your computer skip to the next section.

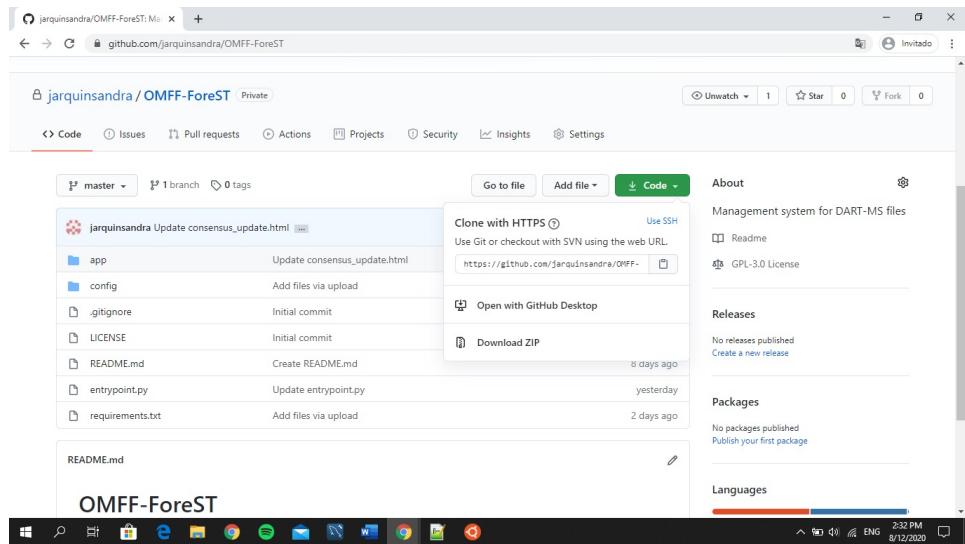
1. Install WLS 1 or 2 depending on your Windows version following the instructions on this website <https://docs.microsoft.com/en-us/windows/wsl/install-win10>. Choose the Ubuntu distribution on the Microsoft Store.

1. FOREST MANAGER ARCHITECTURE



2. Once you have downloaded Ubuntu, click on the icon to start the program, the first time you open it, it will ask you for a username and a password, these two are very important for working on Ubuntu as administrator so you should remember them.

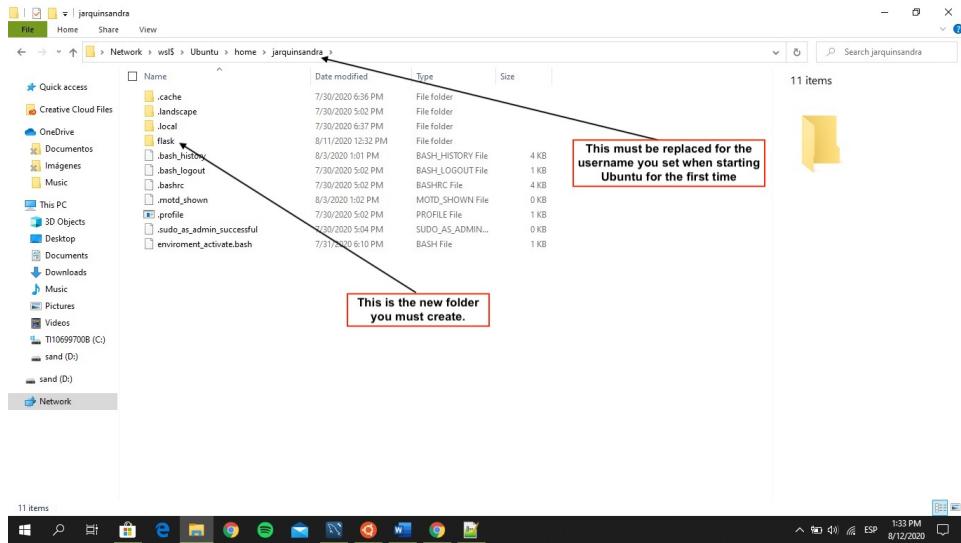
3. Download app from <https://github.com/jarquinsandra/OMFF-ForeST> repository and unzip the content.



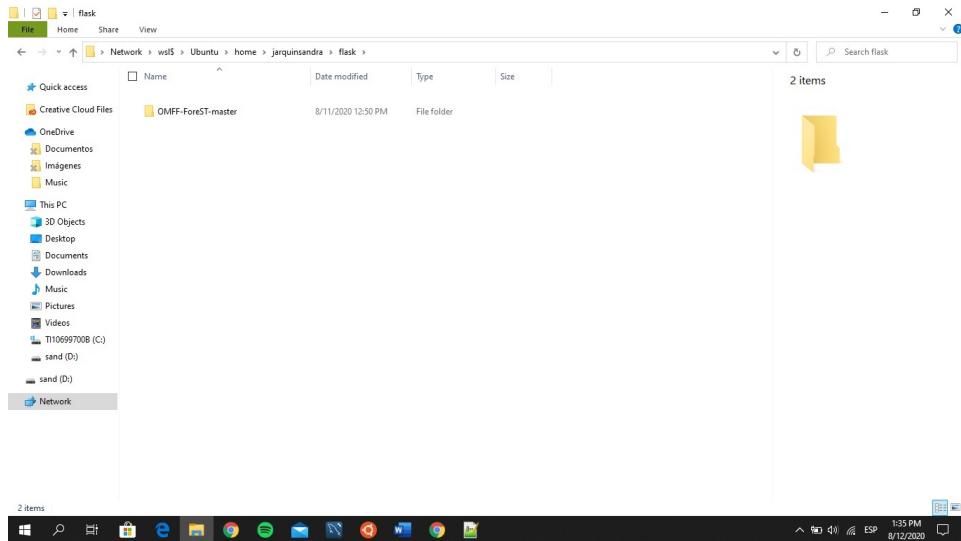
4. Go in the File explorer to the Ubuntu folder for the user you just created, user folder usually is in `\$wsl\Ubuntu\home\username` (this is the username you put when you started Ubuntu), once you are there, create a new folder called flask.

1.1. INSTALLATION ON WINDOWS 10

5



5. Cut and paste the folder OMFF-ForeST-master you downloaded in step 3 in the flask folder you just created.



6. In the Ubuntu terminal go to the folder OMFF-ForeST-master with the following command:

```
cd flask/OMFF-ForeST-master
```

```

● jarquinandre@lvan: ~
Welcome to Ubuntu 20.04 LTS (GNU/Linux 4.4.0-18362-Microsoft x86_64)

 * Documentation: https://help.ubuntu.com
 * Upgrade: https://ubuntuforums.org/advantage
 * Support: https://ubuntuforums.org/advantage

System information as of Wed Aug 12 13:42:52 CDT 2020

System load: 0.32
Usage of /home: unknown
Memory usage: 685
Swap usage: 0
Processes: 7
Tasks: 0
CPU usage: 0.02
IPv4 address for wif10: 192.168.1.139
IPv4 address for wif10: fd04:f1f1:c784:1000:2078:c5d0:608a:4899
IPv4 address for wif10: 2806:102e:b1f8:2:1
IPv4 address for wif10: 2806:102e:b1f8:2:2
IPv4 address for wif10: 2806:102e:b1f8:2:3
IPv4 address for wif10: 2806:102e:b1f8:2:4
IPv4 address for wif10: fd04:f1f1:c784:1000:7000:2d40:ab11:c27a
IPv4 address for wif10: 2806:102e:b1f8:2:5
IPv4 address for wif10: fd04:f1f1:c784:1000:7000:2d40:ab11:c27b
IPv4 address for wif10: fd04:f1f1:c784:1000:7000:2d40:ab11:c27c
IPv4 address for wif10: fd04:f1f1:c784:1000:7000:2d40:ab11:c27d

0 updates can be installed immediately.
13 of these updates are security updates.
To see these additional updates run: apt list --upgradable

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

This message is shown once every day. To disable it please create the
file /etc/periodic/daily/no-hushlogin file.
jarquinandre@lvan: ~

```

7. Install pip and virtual environments in Ubuntu:

```

sudo apt install python3-pip
sudo apt-get install python3-venv

```

This installs the pip installation tool and the virtual environment for all of them. Be aware it will require the password you used to create your Ubuntu account.

8. Create and enter virtual environment called omff, you can name your environment anything you want, but you have to use that name instead of omff when creating and activating the environment.

```

python3 -m venv .omff
source .omff/bin/activate

```

9. Install dependencies, you can install directly from requirements.txt with:

```

pip3 install requirements.txt

```

If the command above produces an error you can run the following commands one by one:

```

pip3 install requirements.txt

```

```

pip3 install Flask
pip3 install Flask-WTF
pip3 install waitress
pip3 install Flask-migrate
pip3 install PyMySQL
pip3 install bokeh
pip3 install Flask-SQLAlchemy
pip3 install pandas

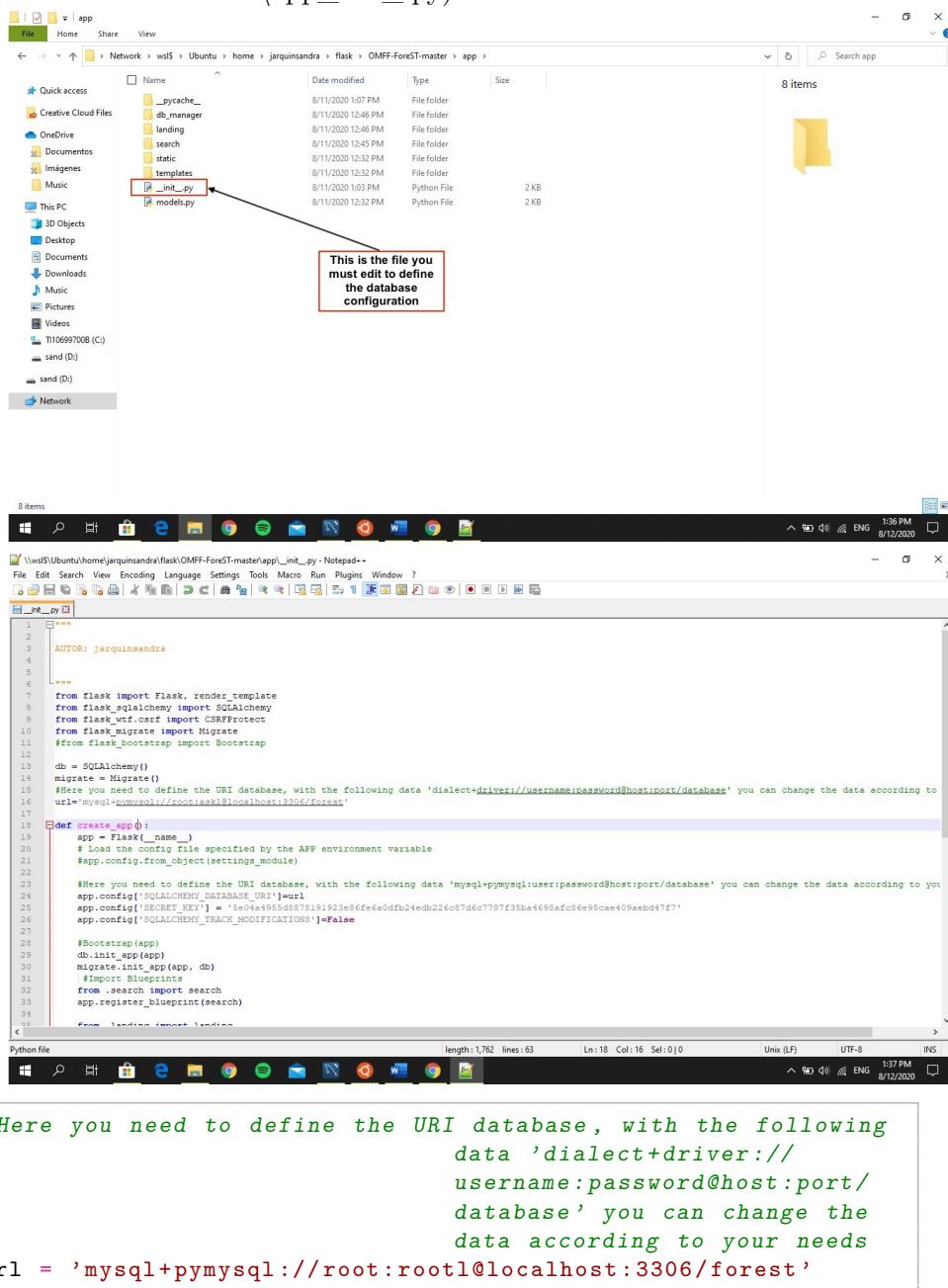
```

10. Download Microsoft Visual Studio. <https://visualstudio.microsoft.com/es/>

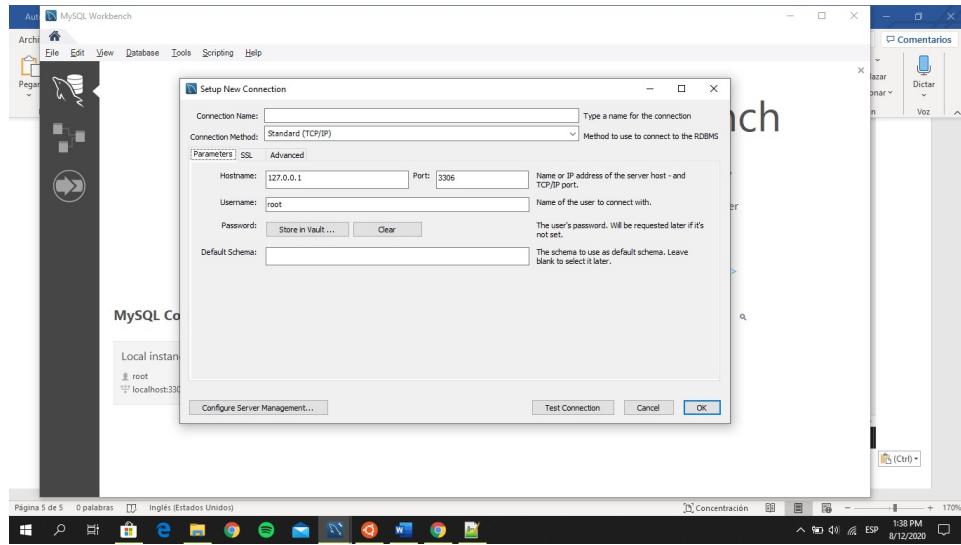
11. Download Mysqlworkbench, be sure to install Mysql server and Mysql workbench, we will use this to build the database.

<https://www.mysql.com/products/workbench/>

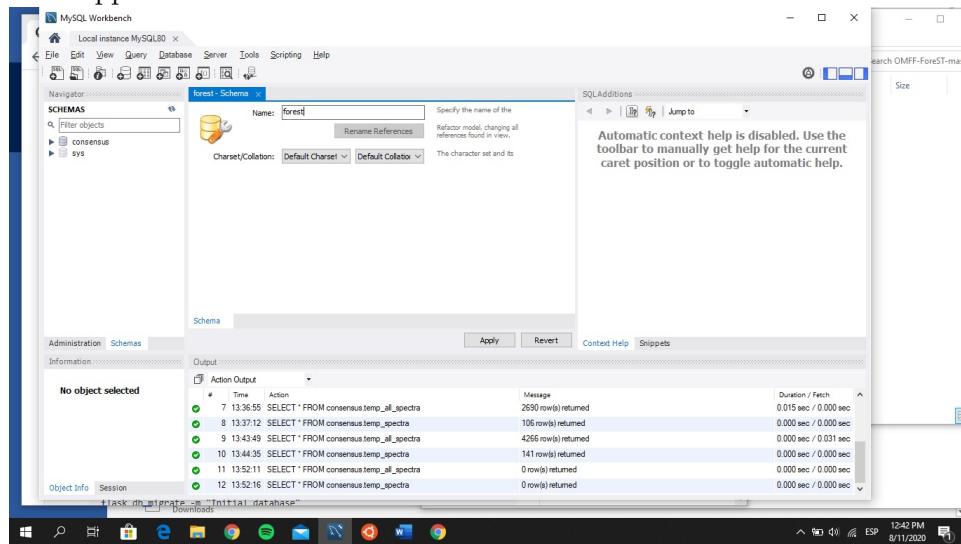
12. Make sure the database is correctly connected to the database, the app is set to work with the database forest, password root and username root, which is contained in the url variable in the init.py file (line 15) inside the app folder, you can open the file in any text editor and change as needed (OMFF-ForeST-master\app_init_.py):



1. FOREST MANAGER ARCHITECTURE



13. Open MySQL Workbench and create a new schema (database) called forest(the name of the database must be the same as the one used in the step number 12). The default user name and password are root, but you can use an specific name and password, just use the port 36600 or be aware of the port you are using for your mysql conection, this information is the one you put in the URI refered in the previous step to setup the database with the app.



14. Initialize the DB with Flask-migrations, run the following in the ubuntu terminal. All changes in the models are given by Flask-migrate, when you run this instructions a new folder called migrations will appear in the OMFF-ForeST folder.

```
flask db init
flask db migrate -m "Initial database"
flask db upgrade
```

15. To setup your computer as a local host. <https://helpdeskgeek.com/windows-10/install-and-setup-a-website-in-iis-on-windows-10/> 15.1 Start typing “turn on windows” in the Search bar. The Turn Windows features on or off utility will show as a result. Click on it.

15.2 The Windows Features window will open. It may take a bit for the different features to load. Once it does, click on the checkbox next to Internet Information Services and then click the OK button.

15.3 The installation will begin and can take several minutes. Once it's completed, click on the Close button.

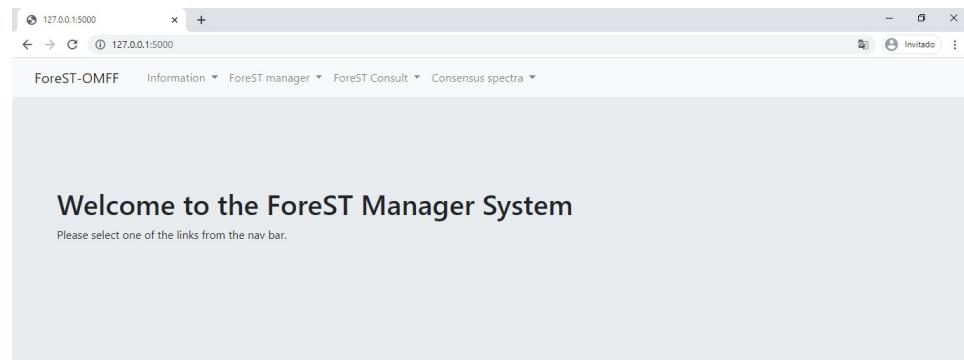
15.4 Go to your browser and type 127.0.0.1 in the direction bar.

16. For this step your environment should be active (if its active you must see your username in the command line, write the following command in the Ubuntu console:

```
python3 entrypoint.py
```

Open web browser and go to 127.0.0.1:5000

The program start page should display.



If you have reached this point successfully, you can start using the OMFF-ForeST spectra manager, if you find any bug or issue please contact me (jarquinsandra@gmail.com) preferably with a screen caption and the description of the problem.

1.1.1 Instructions once you have installed everything in the previous section but you want to run the application again.

This section describes how to run the application if you have followed the last section instructions once in your computer.

1. Open Ubuntu.
2. Go to the application folder in the terminal:

```
cd flask/OMFF-ForeST-master
```

3. Start the virtual environment:

```
source .omff/bin/activate
```

4. Open MySQLWorkbench and start your connection
5. Start the application:

```
python3 entrypoint.py
```

6. Open a web browser and go to 127.0.0.1:5000, application
7. If you finished using the app you should stop the server (Ctrl + C), exit the virtual environment and close Ubuntu and MySQLWorkbench.

```
#Exit virtual environment
deactivate
```

1.1.2 Database

ForeST Manager was implemented with a MYSQL database, however since it uses SQLAlchemy, the database can be easily changed to other relational DB Management system or a non Relational DB system. The database is organized in four tables as follows:

- files table: this contains all the information about the input data, including file name, species (in the format GenusSpecies), WD number, mz, intensity, time-created and time-updated.

The screenshot shows the MySQL Workbench interface with the 'files' table selected. The results grid displays 14 rows of data, each representing a file entry with columns: id, filename, intensity, mz, origin, species, wd, time_created, and time_updated. The data includes various file names like 'AbiesAba_WD164963_MADw1399.txt' and their corresponding metadata. The bottom pane shows the table definition with columns: id (int AI PK), filename (varchar(11)), intensity (float), mz (float), origin (varchar(11)), species (varchar(6)), and wd (varchar(6)).

id	filename	intensity	mz	origin	species	wd	time_created
464	AbiesAba_WD164963_MADw1399.txt	0.3	95.059	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
465	AbiesAba_WD164963_MADw1399.txt	10.4	96.05	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
466	AbiesAba_WD164963_MADw1399.txt	1.1	96.10	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
467	AbiesAba_WD164963_MADw1399.txt	0.1	95.194	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
468	AbiesAba_WD164963_MADw1399.txt	4.1	97.036	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
469	AbiesAba_WD164963_MADw1399.txt	0.2	97.141	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
470	AbiesAba_WD164963_MADw1399.txt	0.2	98.04	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
471	AbiesAba_WD164963_MADw1399.txt	0.2	98.066	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15
472	AbiesAba_WD164963_MADw1399.txt	1.7	99.05	MADw1399	AbiesAba	WD164963	2020-08-11 13:15:15

- reference-spectra: this table contains all the consensus spectra for each one of the species in the database. This table is updated everytime the update consensus button is clicked in the app.

The screenshot shows the MySQL Workbench interface with the 'reference_spectra' table selected. The results grid displays 9 rows of data, each representing a consensus spectrum entry with columns: id, species, mz, int_rel, and time_created. The data includes various species names like 'AbiesAba' and their corresponding mz values. The bottom pane shows the table definition with columns: id (int AI PK), species (varchar(11)), mz (float), int_rel (float), and time_created (datetime).

id	species	mz	int_rel	time_created
1	AbiesAba	96.0485	21.2533	2020-08-11 13:25:59
2	AbiesAba	97.034	15.4792	2020-08-11 13:25:59
3	AbiesAba	98.0375	14.9722	2020-08-11 13:25:59
4	AbiesAba	99.047	5.6316	2020-08-11 13:25:59
5	AbiesAba	100.0475	1.1143	2020-08-11 13:25:59
6	AbiesAba	106.079	1.11667	2020-08-11 13:25:59
7	AbiesAba	106.09	1.1875	2020-08-11 13:25:59
8	AbiesAba	108.085	1.05	2020-08-11 13:25:59
9	AbiesAba	109.033	3.19286	2020-08-11 13:25:59

- temp_all_spectra : this table is temporal for the calculation of custome spectra, it is filled with the spectra from the files submited and erased each time this section is used.
- temp_spectra: this table is also temporal for storing and showing the custome consensus spectra, it is filled ans erased each time the section its used.

To connect the database with the app SQLAlchemy was used, models where created for the four tables, and everything was setup to work with Flask-migrations control, tables maybe added to the program altering the script and upgrading the models through this package directly on the terminal.

Each time a table is change or included it should be done using migrations with the following commands directly on the terminal:

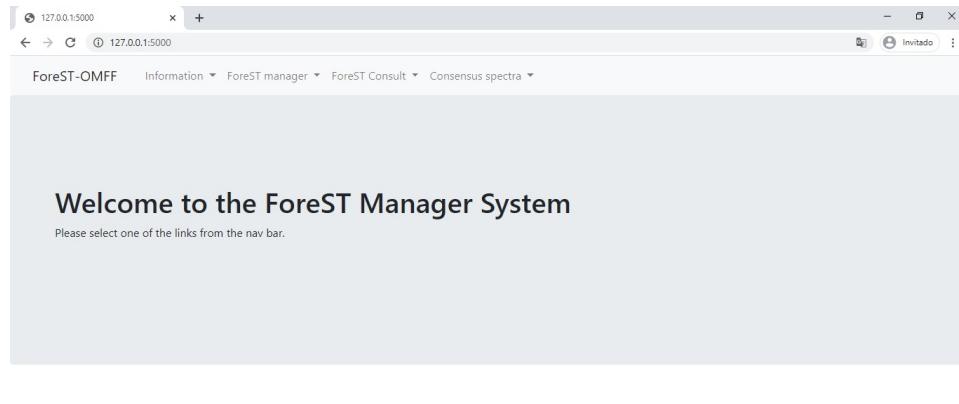
```
flask db migrate -m "message to identify the changes in the DB"  
flask db upgrade
```

MySQLWorkbench may be used to visualize the database, however it is not recommended to execute changes directly here, it must lead to problems in the synchronization of the application.

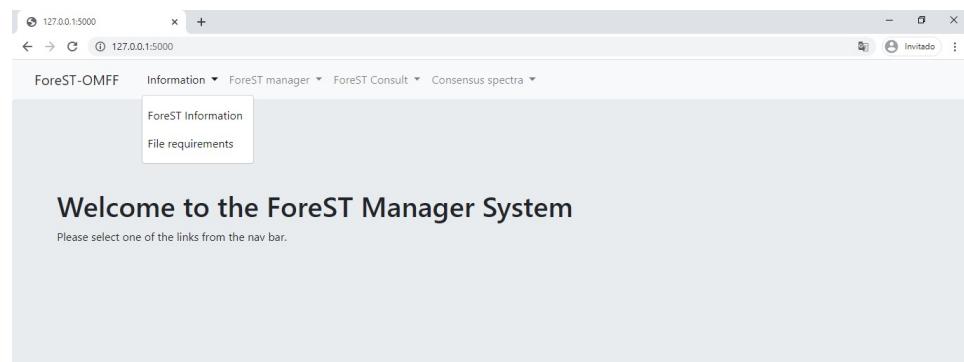
2

OMFF-ForeST Manager interface

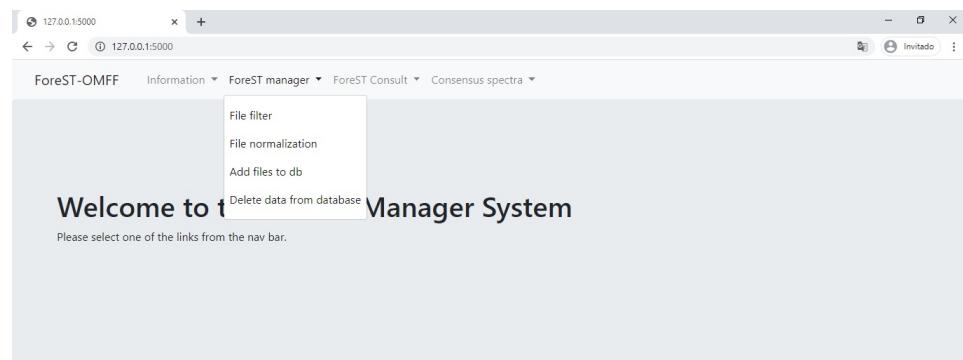
The OMFF (Open Mass Fingerprinting Framework) ForeST Manager application interfaced is displayed on a web browser in the address 127.0.0.1:5000, if you installed everything correctly according to the instructions in the chapter 1, shoul should look something like this:



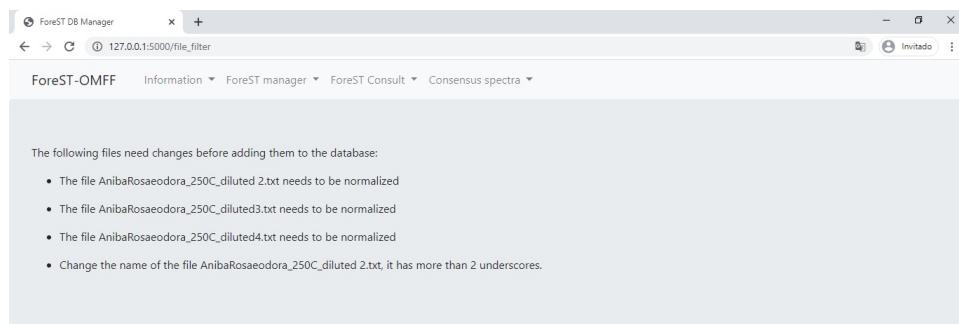
The application includes four options in the navigation bar, they functions of each one will be explained here. The first one is information. Here you can find the information about the application and the database, and also the requirements to include files in the ForeST.



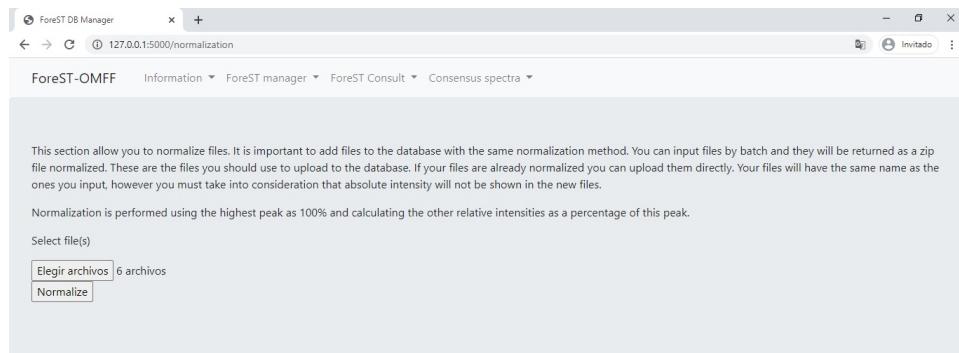
2.1 Database manager



- File filter: this module should be used to identify problems in your files prior to their inclusion in the database, it recognizes spaces, the presence of more underscores than allowed and non normalized files. If the files you input in the filter are correct, no message or file will appear.



- File normalization: here you can normalize your files, the normalization is performed based on the maximum intensity peak detected as 100% and the rest of them as a percentage of this peak. The non normalized data will be lost in the new files, you can use several files as input and the output will be a zip file containing all the normalized intensity files with the same name as the input file.



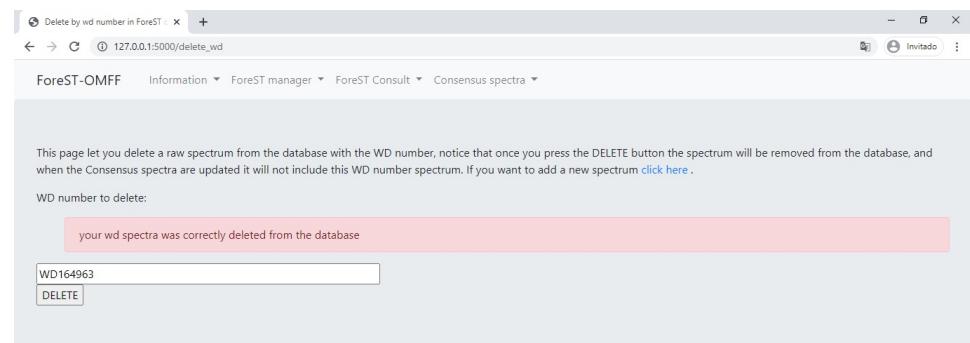
- Add files to db: in order for you to add files to the ForeST database, you need to pass them through the file filter first, and correct the errors, once you have done this, you can add your text files to the database, they must contain in the first column the m/z value, and in the second column the relative intensity value, any other information the file contains will not be included in the database. The application uses

the name of the file to obtain the information about the sample, so you must follow the requirements.

Once your files have been uploaded to the database, a message will appear in the web page, please do not refresh or cancel the process, this could cause errors.

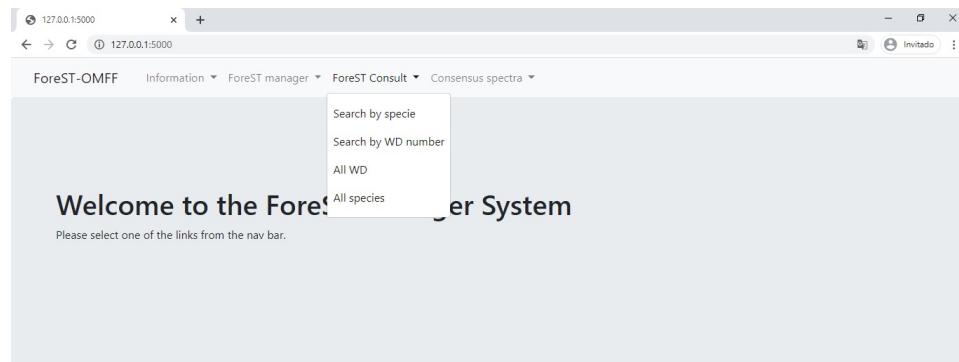


- Delete data from Database: This feature allow you to delete a complete spectra from the database with its WD number, it is very important that you are sure you want to delete this spectra, otherwise you will have to add it again in the module add files to the db.

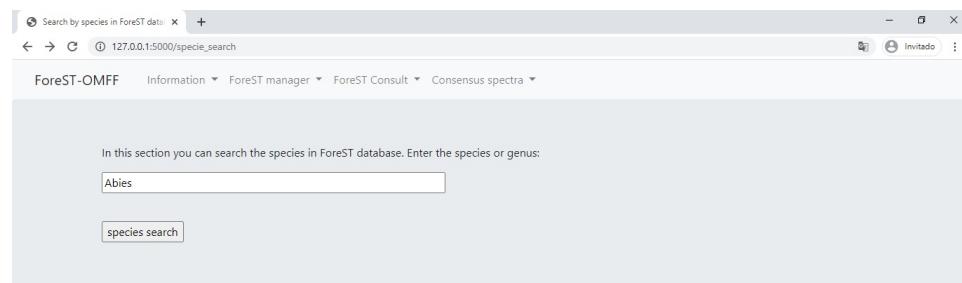


2.2 ForeST consult

You can do searches directly in the forest database in MySQLWorkbench, however this program let you do simple searches in the database.



- Search by species: to search a genus or specie in ForeST you just need to type the word you want to match against the database, this search is not case sensitive, In the shown example the search shows all entries corresponding to the genus Abies.



The screenshot shows a web browser window titled "Search by species in ForeST data" with the URL "127.0.0.1:5000/specie_search/Abies". The page displays a table of search results for the query "Abies". The table has columns: filename, species, wd, origin, and time_created. The results are:

	filename	species	wd	origin	time_created
0	AbiesAlba_WD164964_MADw1400.txt	abiesalba	WD164964	MADw1400	2020-08-11 13:13:28
712	AbiesAlba_WD164965_MADw1401.txt	abiesalba	WD164965	MADw1401	2020-08-11 13:13:28
1719	AbiesAlba_WD164966_MADw2444.txt	abiesalba	WD164966	MADw2444	2020-08-11 13:13:28
2558	AbiesAlba_WD164962_MADw10537.txt	abiesalba	WD164962	MADw10537	2020-08-11 13:16:34
3021	AbiesAlba_WD164973_MADw8543.txt	abiesalba	WD164973	MADw8543	2020-08-11 13:16:34

- Search by WD number: this search gives you the information of a single spectra which is identified by its WD number.

The screenshot shows a web browser window titled "Search by wd number in ForeST" with the URL "127.0.0.1:5000/wd_search". The page contains a text input field with "WD164964" and a button labeled "WD search". A message above the input field reads: "This search will display the information available in ForeST database for a WD number. Write in the box the wd number you want to search. You should include wd followed by the number, for example WD66432."



The screenshot shows a web browser window titled "Search by wd number in ForeST". The URL is 127.0.0.1:5000/wd_search/WD164964. The page header includes "ForeST-OMFF", "Information", "ForeST manager", "ForeST Consult", and "Consensus spectra". Below the header, a message says "The results for your search on WD164964 number are:" and "The first column is an ID number for the MYSQL database and is does not reflect any specific order." A table follows:

	filename	species	wd	origin	time_created
0	AbiesAlba_WD164964_MADw1400.txt	AbiesAlba	wd164964	MADw1400	2020-08-11 13:13:28



- All WD: this search takes a while to load, it gives you information about all the spectra in the database, and you can download the results as a csv file.

The screenshot shows a web browser window titled "All WD numbers with their specie". The URL is 127.0.0.1:5000/wd_all. The page header is identical to the previous screenshot. Below the header, a message says "These are all the WD numbers in ForeST database." A "Download csv" button is visible. A table follows:

	species	wd
0	AbiesAlba	WD164964
712	AbiesAlba	WD164965
1719	AbiesAlba	WD164966
2558	AbiesAlba	WD164962
3021	AbiesAlba	WD164973
3562	AbiesAlba	WD165569

- All species: this section present you a summary of the number of spectra by species in the database, this section might take a while to load, and the results can be downloaded as a csv file.

	species	spectra_number
0	AbiesAlba	10
1	AbiesAmabilis	19
2	AbiesBalsamea	25



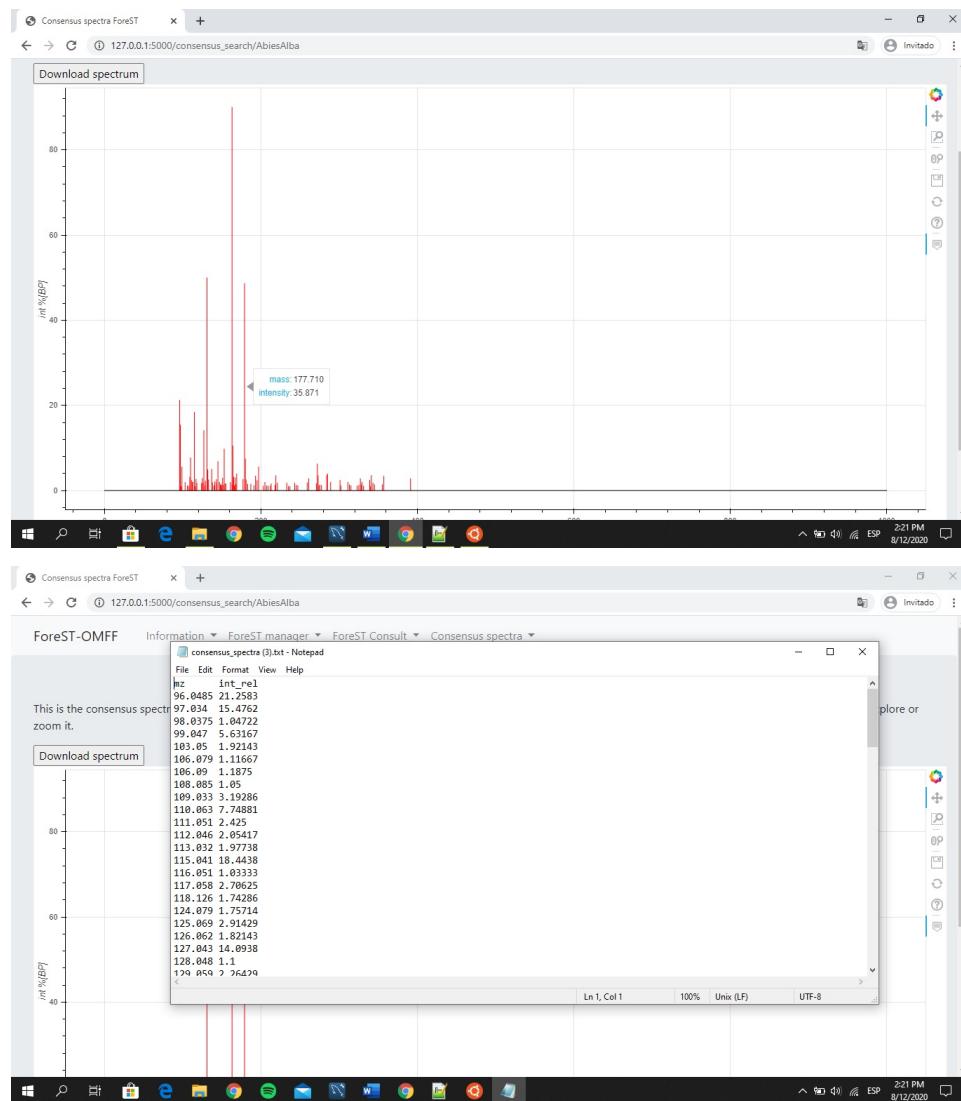
2.3 Consensus spectra

A consensus spectra is the spectra calculated using all the available spectrum for a species. All consensus spectrum are calculated using a Sliding Window Algorithm approach. In order to calculate the consensus spectrum of a class the peaks from all the spectra for the class must be binned to make them comparable. The sliding window algorithm idea is to replace the original m/z values of all peaks with a set of calibrated m/z values. The algorithm is initiated by calculating, at each m/z point, the total number of peaks, in all samples, that are within the window of potential shift for the m/z point. The m/z point that has the highest total number of peaks (summing over all samples for each class) within its window of potential shift is entered into the new m/z set as a calibrated m/z value, and all the points that are within the window of potential shift for this point are removed from the subsequent steps of the algorithm.

Then, the above procedure is repeated (ie, finding the point, from the remaining points, that has the highest total number of peaks within its window of potential shift, entering it into the new mass/charge set, and removing the mass/charge points that are within its window of potential shift from the subsequent steps of the algorithm) until all peaks are exhausted from every sample.

The end product of this repeated procedure is the new m/z set or consensus spectra. The final step of the algorithm is to construct a calibrated data set that consists of m/z and average intensity for each species, recognizing as peaks only those who are present in a certain percentage of samples, this algorithm helps recognize the peaks that are useful for classification purposes, redundant detection allow the components to be confidently reported.

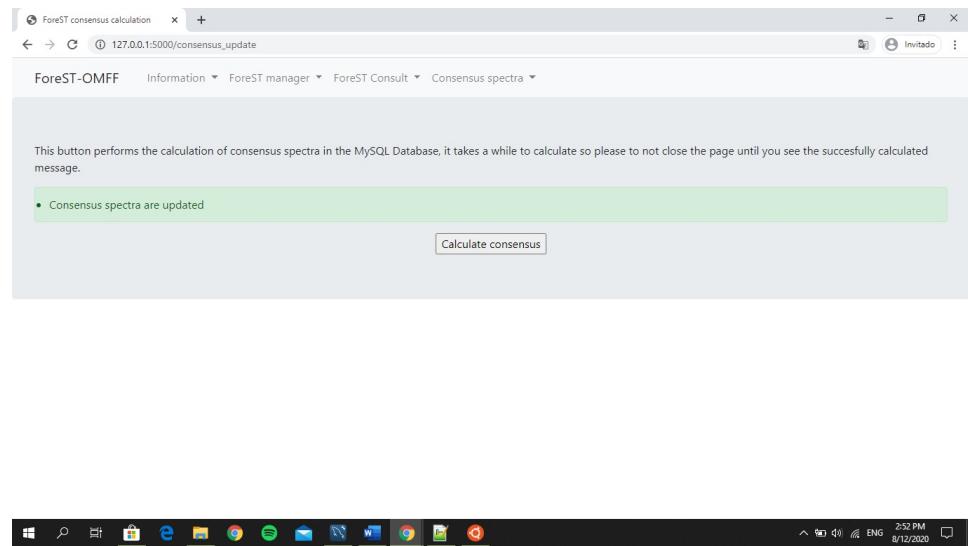
- Search consensus spectra: here you can search, visualize and download a consensus spectra that is already in the ForeST database, for this to happen you need to update the consensus spectra at least once. Tools in the right side of the image are useful to explore the spectra and download it as image (PNG).



- Update consensus spectra: this step is for updating the consensus spectra in the database, this procedure must be performed one or twice a month, or each time an update to the search tool for comparing spectra needs to be updated (Arbor Harbor web page for ForeST search). You need to press the button calculate consensus one time, and wait for the calculation to be performed, note that the calculation may take a while and you must not interrupt the process or turn off the computer

or MySQLWorkbench server, this may lead to errors in calculations.

Once the message in green "Consensus spectra are updated" appears you can continue using the application.



- Consensus calculation: here you can calculate customized consensus spectra uploading txt files of the spectra you want to use to calculate the consensus spectrum. The spectrum is displayed and can be explored with the tools on the right side of the screen. The default parameters are set, but you can play around putting different parameters.

The consensus spectrum is calculated using an sliding window algorithm to determine the real peaks for all the spectra. The size of the bins used for the calculation is 0.003 m/z units, and the window size is 0.01. All peaks below 1% (relative intensity) are considered to be noise and the peak presence of 0.5 means that only peaks present in at least 50% of the samples are taken into account to calculate the consensus spectra. You can manipulate these parameters indicating different values in the corresponding box.

ForeST-OMFF Information ▾ ForeST manager ▾ ForeST Consult ▾ Consensus spectra ▾

Here you can calculate the consensus spectrum for a group of txt files. All the files must be normalized, if they are not you can [click here](#) to normalize them. They should contain two columns the first with m/z values and the second with the relative intensity.

The consensus spectrum is calculated using a sliding window algorithm to determine the real peaks for all the spectra. The size of the bins used for the calculation is 0.003 m/z units, and the window size is 0.01. All peaks below 1% (relative intensity) are considered to be noise and the peak presence of 0.5 means that only peaks present in at least 50% of the samples are taken into account to calculate the consensus spectra. You can manipulate these parameters indicating different values in the corresponding box.

Select file(s) to calculate the consensus spectrum:

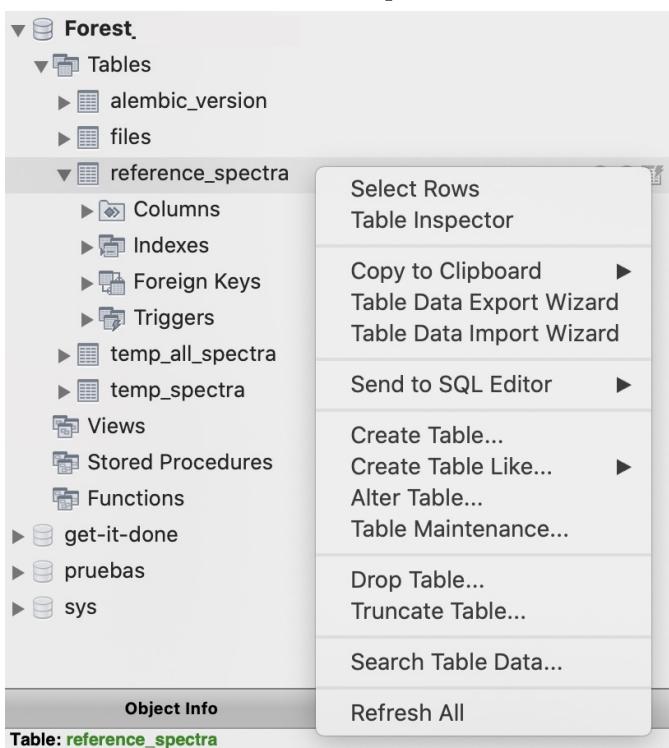
no files selected

bin	0.003
window	0.01
noise level	1.0
peak presence	0.5

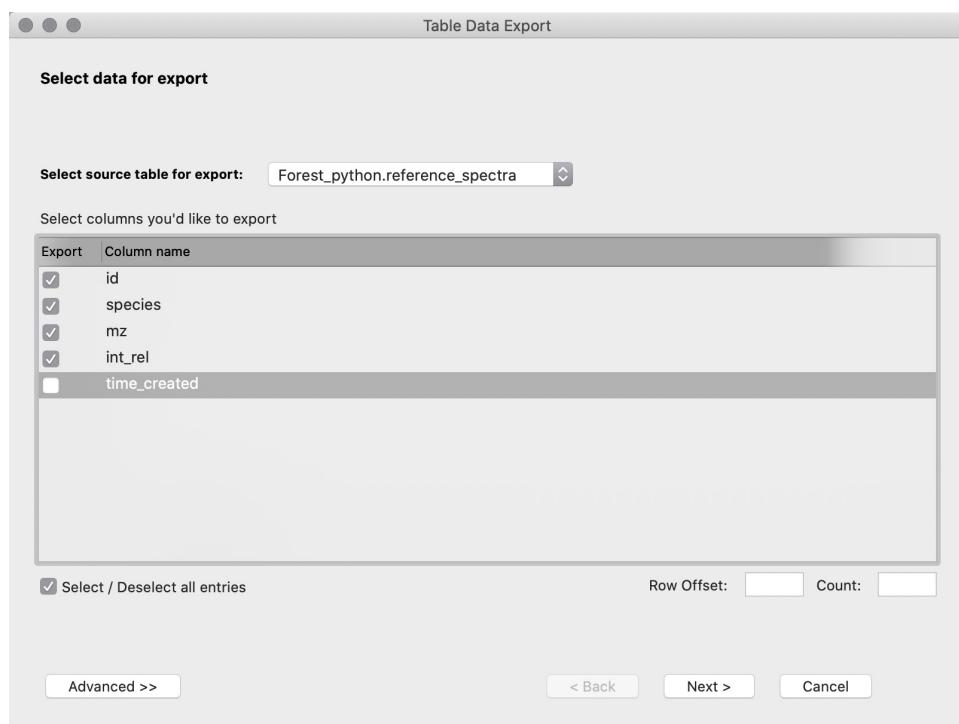
3

ForeST web page update for Arbor Harbor

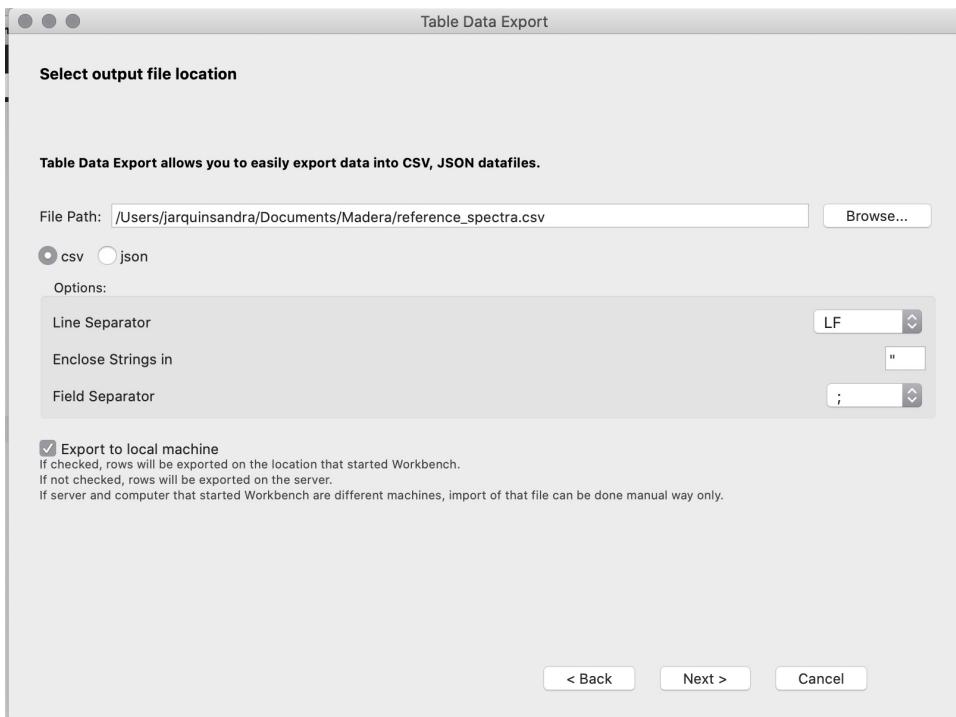
ForeST web page was designed with PHP and uses a MySQL database, the database has only one table corresponding to the consensus spectra calculated in this application. The table should be extracted from MySQLWorkbench. If you are working with the application, MySQL Workbench must be open and working as local server, go to the forest database (where all our tables from the application are stored), left click on the table reference spectra and select Table Data Export Wizard.



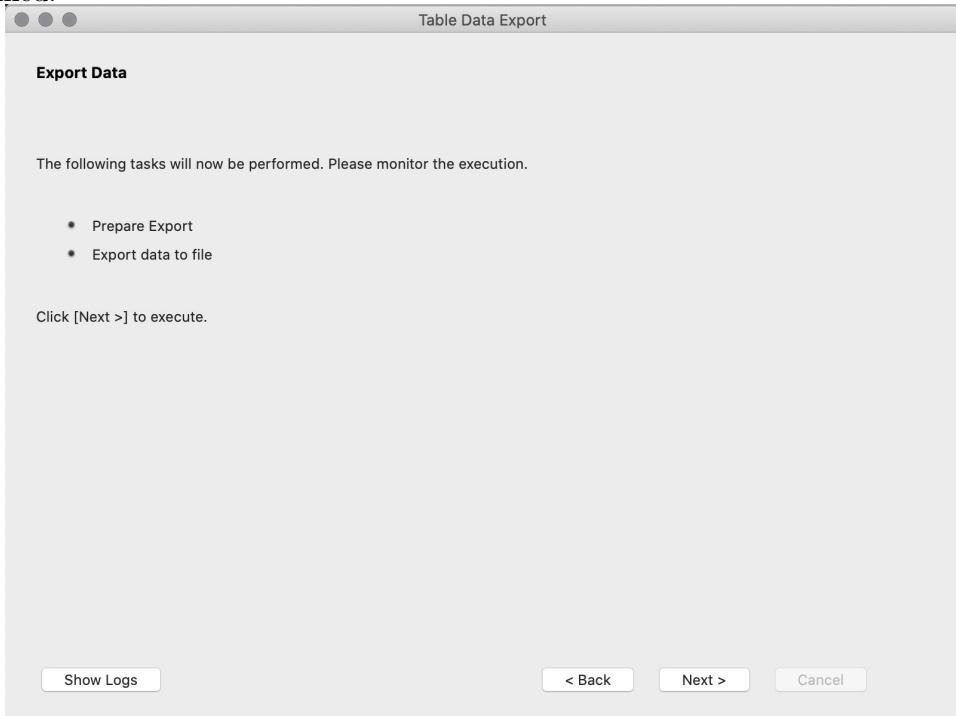
The Table Data Export Will appear, and you must deselect the box for the last field time_created and then press next.



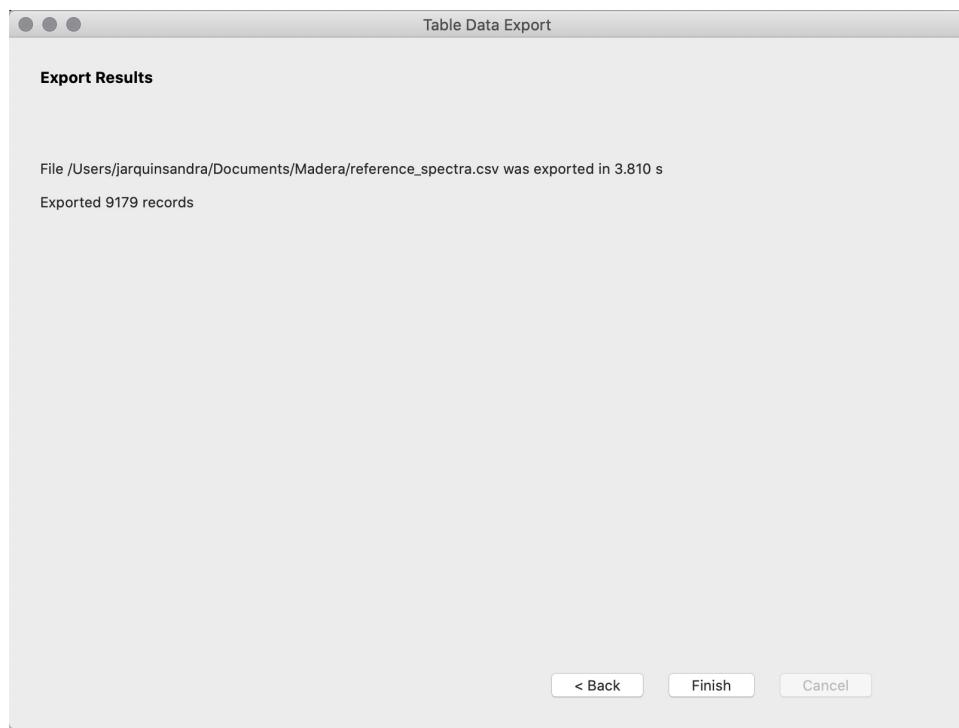
The next screen will give you the option to select the format for the table export, here we must use the csv format and select the path where the csv file is stored in your machine.



Next screen just gives you a summary of the process that will be performed.



And finally just finish the process, now your file will be in the path you selected as csv file.



This file must be given to the administrators of Arbor for its inclusion in the web page. The file should replace the content of the table spectra_web in the ForeST web page.

If you have any trouble or error with the process, you can reach me at jarquinsandra@gmail.com