



MASTER OF SCIENCE  
IN ENGINEERING

UNIVERSITY OF APPLIED SCIENCES WESTERN SWITZERLAND  
MSE - SOFTWARE ENGINEERING

---

## Projet d'approfondissement Inphinity

---

*Authors:*  
Déruaz Vincent

*Supervisors:*  
Prof. Carlos Peña

**Hes·SO**

Haute Ecole Spécialisée  
de Suisse occidentale

Fachhochschule Westschweiz

University of Applied Sciences and Arts  
Western Switzerland

Lausanne, June 4, 2016

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	foreword . . . . .	3
1.2	phages Vs bacteria . . . . .	3
<b>2</b>	<b>State of the art</b>	<b>4</b>
2.1	Biology . . . . .	4
2.2	Bio-informatic . . . . .	4
<b>3</b>	<b>Methods</b>	<b>5</b>
3.1	Docker . . . . .	5
3.1.1	Functionment . . . . .	5
3.1.2	Docker-machine installation . . . . .	5
3.1.3	Docker commands . . . . .	6
3.1.4	Inphinity, build & run . . . . .	6
3.2	Phamerator . . . . .	7
3.2.1	Installation . . . . .	7
3.2.2	How it's works . . . . .	8
3.3	PhamDB . . . . .	9
3.3.1	Installation & Run . . . . .	9
3.3.2	Utilisation . . . . .	10
3.4	Database & Dataset . . . . .	11
3.5	Phamily . . . . .	12
3.5.1	Alignment . . . . .	12
3.5.2	Phages selection . . . . .	12
3.5.3	Data completion . . . . .	12
<b>4</b>	<b>Results &amp; Analyse</b>	<b>13</b>
4.1	First result . . . . .	13
4.1.1	Tree . . . . .	13
4.1.2	Hosts . . . . .	13
4.2	Database . . . . .	13
4.2.1	SEA . . . . .	13
4.2.2	Phages list integration . . . . .	13
<b>5</b>	<b>Conclusion</b>	<b>14</b>
5.1	Problems encountered . . . . .	14
5.1.1	Phamerator Installation . . . . .	14
5.1.2	SEA database . . . . .	14
5.1.3	PhamDB Limitation jobs . . . . .	14
5.1.4	PhageDB.org . . . . .	14
5.2	Perspectives . . . . .	14
5.2.1	Database population . . . . .	14
5.2.2	Resultats validation . . . . .	14
5.2.3	Results by host . . . . .	14
<b>6</b>	<b>References</b>	<b>14</b>



# 1 Introduction

## 1.1 foreword

This project falls within the context of a thesis proposed by Prof. Carlos Peña, YokAi Que, MDPhD and Grégory Resch, PhD entitled *In silico prediction of phagebacteria infection networks as a tool to implement personalized phage therapy* [2].

The official statment of the project is:

By using automated learning methods, explore alternate metodologies for bacteria and phages interaction modelisation on genomic informations or proteins.

## 1.2 phages Vs bacteria

In our modern world a challenging issues has apear concerning conventional antibiotics. In deed, some batceria have developpe resistance to antibiotics. To overcome this people are looking at phage therapy.

Phage therapy is the utilisation of phages, bacteriophage viruses, to threat infectious diseases of bacterial origin. This therapy is known to have only very few and only benign side effets. This last point make phagotherapy, not only useful to avoid antibiotic in case of resistance, but also to avoid their "toxicity".

Briefly, phage therapy was the only threatment solution in the before the 1930's. The apearence of the penicillin in the early 1940's and other modern drugs, releagate phage therapy to the past. But, in the slavic countries, phage therapy continued to be used as a current treatment.

Luckly for us, we don't have to start from nothing in phage therapy. However, we have the necessity to find a way, a methode to validate the phage selection. [3]

## 2 State of the art

### 2.1 Biology

TODO: genome

### 2.2 Bio-informatic

TODO: sequence alignment

TODO: explain phamily

TODO: explain genebank

## 3 Methods

In this section we will discuss about what we've used during this project. The Docker technology is used to build the different work environment. Phamator and PhamDB are used to compute genomes into family. Everything is stored into some SQL databases.

### 3.1 Docker

Docker allows to package applications with a functional system and every dependencies needed to run it, into a standardized container. [1]

#### 3.1.1 Functionment

In a specific file called 'Dockerfile' you describe a system. You can build and run this system everywhere docker engine is installed. It will create a container, containing your application.

Containers are an isolated system from host or other containers. You can use every Linux distribution to run your container.

Docker builds images using layers, it allows docker to share those layers between containers, reducing disk usage at best.

#### 3.1.2 Docker-machine installation

You have to install docker on your system, it can be used on Linux, Windows or MacOS.  
<https://docs.docker.com/linux/>

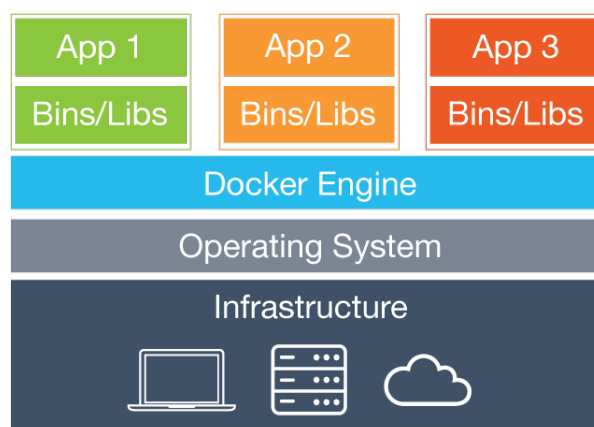


Figure 3.1: Docker harchitecture

## Windows

If you want to use docker on Windows you need to do as following to ensure to have a docker-machine with more than 20Go.

```
$ docker-machine rm default
$ docker-machine create -d virtualbox --virtualbox-memory=4096 --virtualbox-cpu-count=2
  → --virtualbox-disk-size=50000 default
```

It will crate a docker-machine with 50Go of disk space.

### 3.1.3 Docker commands

Here are the basic commands you will need to manage docker. Attention, you will need to be in the same directory as the Dockerfile.

This command allow to build an image describe in the Dockerfile.

```
$ docker build -t "<image Name>" .
```

This command is used to run a container using a pre-build image, with a binding port.

```
$ docker run -it -d -p <host port>:<container port> <image name>
```

If you need to acces the container bash console, juste use this commande

```
$ docker exec -i -t <container ID> /bin/bash
```

You can list all the running containers and use a <container id> to stop it.

```
$ docker ps
$ docker stop <container ID>
```

This last command give you the ip of your docker-machine.

```
$ docker-machine ip
```

### 3.1.4 Inphinity, build & run

For the main code of this project we use python through Jupyter. To do so you can find a docker image that run Jupyter, python3 and some machin learning libraries. go to "dockers/jupyter\_align\_mysql" directory.

To build and run the environment type the following commands:

```
$ docker build -t "pa/inphinity" .
$ docker run -v <path to project dir>/jupyter_align_mysql/src:/home/pa/work/ -i -t -d -p 8888:8888
  → pa/inphinity
```

Replace <path to project dir> by the entire path to the directory. If you want to, you can change the host port. Just change "-p 8888:8888" to "-p <any port>:8888".

You can now access the jupyter interface and the project files using any browser you want using `http://<docker-machineip>:8888`

At this point you should see the interface figure 3.2

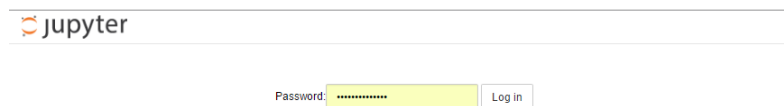


Figure 3.2: Jupyter login page

**Rq:** The password is "Inphinity-more"

When you're logged in you can access the "inphinity" directory. It contains most of this project results. We will discuss them later in this document.

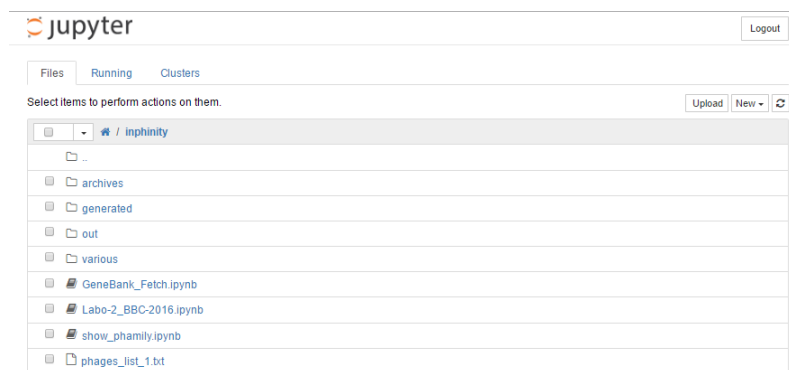


Figure 3.3: Inphinity jupyter directory

## 3.2 Phamerator

Phameroatr is a *bioinformatic tool for comparative bacteriophage genomics* [4]. Phamerator will allow us to compute and store phamily, in a database.

### 3.2.1 Installation

To use phamerator you have to install a linux environment. Thus copy the file "phamerator.sh" from the directory and execute it in order to install and run Phamerator.

Normally you Phamerator should start at the end.



Name	Length (bp)	GC %	Number of Genes	Cluster
Corndog	69777	65.384000	122	
Wildcat	78441	56.851600	148	
Giles	54512	67.267800	79	
Marvin-DRAFT	65089	63.419300	111	
Send513-DRAFT	71547	55.998200	101	
Bxb1	50550	63.649900	86	A1
112	51277	63.699100	81	A1

**The Actinobacteriophage Database**  
at PhagesDB.org

Home Phages Hosts Data BLAST Publications Resources Software Social

Search PhagesDB.org

Recently Added Phages

- BossLady
- Birchlyn
- Attoomi
- Rowa

**Mycobacterium phage Marvin**

Downloading BLAST (0% Complete)

Figure 3.4: phamerator interface

### 3.2.2 How it's works

TODO: explain

Phamily	Number of Members	Clusters
7	26	A1, A2, A3, A4, A5
8	26	A1, A2, A3, A4, A5
9	26	A1, A2, A3, A4, A5
10	26	A1, A2, A3, A4, A5
11	26	A1, A2, A3, A4, A5
17	16	A1, A3, A5
18	11	A1

Gene	Phage	Global % Identity	BLAST E-Value
16	Bxb1		
20	Bxz2		
gp17	U2		
gp17	Bethlehem		
20	D29		
gene20	L5		
22	Che12		
19	DD5		
19	Jasper		
KBG_19	KBG		
18	Lockley		
22	Pukovnik		
18	Solon		
19	Peaches		
20	SkiPole		
23	RedRock		

Downloading BLAST (0% Complete)

Figure 3.5: phamerator interface

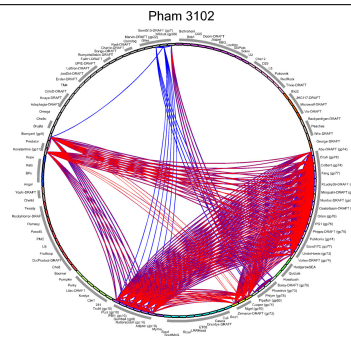


Figure 3.6: phamerator interface

## 3.3 PhamDB

To facilitate the construction of databases containing our phamily we will use PhamDb. In deed, with phamDb we can populate a database with new phage. At every addition of phage, it will recompute phamily accordingly to the new phage added.

We no longer need to access to Phamerator by GUI. In the future it will let us build a fully automated pipline of actions.

### 3.3.1 Installation & Run

To build and run the environment type the following commands:

```
$ docker build -t "pa/phamdb" .
$ docker run -v <path to project dir>/jupyter_align_mysql/src:/home/pa/work/ -i -t -d -p 81:80
  → pa/phamdb
```

Replace <path to project dir> by the entire path to the directory. If you want to, you can change the host port. Just change "-p 81:80" to "-p <any port>:80".

You can now acces the jupyter interface and the project files using any browser you want using <http://<docker-machineip>:80>

At this point you should see the interface figure 3.4, but with no database for the moment.

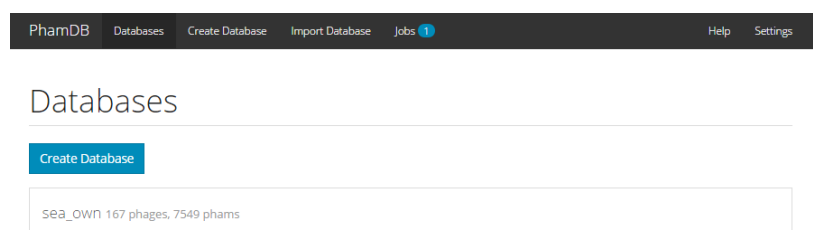


Figure 3.7: Phamdb interface

### 3.3.2 Utilisation

As you see from figure 3.4 you can access the list of all existing database and consult them.

PhamDB Databases Create Database Import Database Jobs 1 Help Settings

## Database - sea\_own

Edit Database Delete database

167 phages, 7549 phams

**Description:** sea\_own first tests  
**Orphans:** 61.5% (4646 / 7549)  
**Conserved domain search:** Yes  
**Download:** [sea\\_own.sql](#)

### Connecting with Phamerator

To view this database in Phamerator, click **Edit** > **Settings**, then add the following settings:

Server	http://home.kamylh.ch:81/db/
Database	sea_own

Phages	
<b>lambda</b> (id: 10710, 73 genes)	Download
<b>NF</b> (id: 10753, 27 genes)	Download
<b>B103</b> (id: 10778, 17 genes)	Download
<b>B55</b> (id: 1126949, 272 genes)	Download
<b>vB_AsaM-56</b> (id: 1127514, 83 genes)	Download
<b>TA17A</b> (id: 1131248, 94 genes)	Download
<b>phage</b> (id: 1131248, 94 genes)	Download

Figure 3.8: PhamDb database visualisation

You can create a new database from three different ways:

1. By importing phages from existing database on phamdb.
2. By uploading Genbank files.
3. By importing as an Sql file.

PhamDB

Databases

Create Database

Import Database

Jobs 1

Help

Settings

## Create Database

Database Name

Enter a name for the database

Description

Enter a description

☐ Search Conserved Domain Database

Search for each gene in NCBI's conserved domain database. This will significantly increase runtime.

## Import Phages from Database

Select a phage to import it from a database.

Databases

sea\_own (id: 157 phages)

Phages

lambda (id: 10710, 73 genes)

**NF** (id: 10753, 27 genes)

**B103** (id: 10778, 17 genes)

B55 (id: 1126949, 272 genes)

vB\_AsaM-56 (id: 1127514, 63 genes)

TA17A (id: 1131248, 94 genes)

☒ **B103** (id: 10778, database: sea\_own, 17 genes)

☒ **NF** (id: 10753, database: sea\_own, 27 genes)

## Upload Genbank Files

Sélectionner fichiers

Aucun fichier choisi

Select genbank files or drag and drop.

## Import Database from SQL File

Database Name

Enter a name for the database

Description

Enter a description

## Upload SQL File

Choisissez un fichier

Aucun fichier choisi

No file selected.

Submit

Figure 3.9: PhamDb database creation

### 3.4 Database & Dataset

We use the default database from phamerator for this phase of the project in order to gain some time. You can see the database structure on the figure 3.5 .

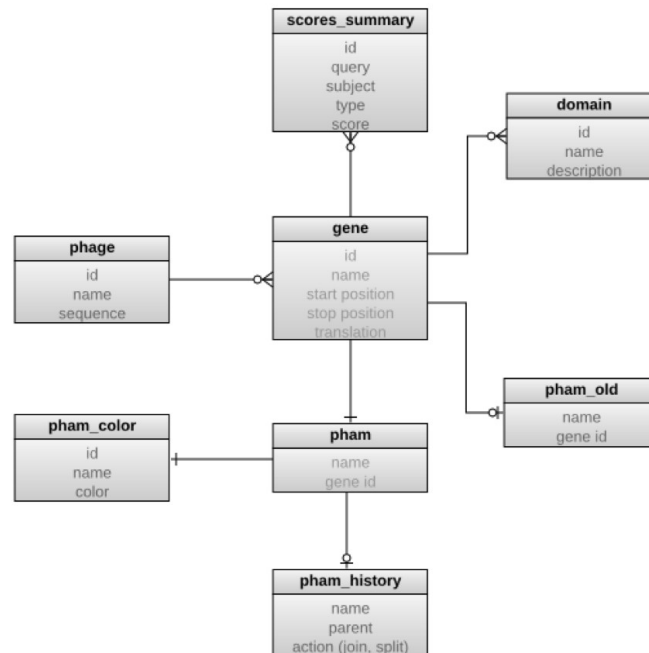


Figure 3.10: Docker harchitecture

From the dataset at my disposal, I've used the list of phage "phages\_list\_1.txt". You can find it in annexe of this document.

## 3.5 Phamily

Now we will discuss about phamily and the main script realise during this phase of the project. You can see and run the folowing code from the jupyter interface, Cf. chapter 3.1.4.

### 3.5.1 Alignment

**Have to:** You are now in your browser and have open the source file "inphinity/show\_phamily.ipynb". The class "sea\_inphinity()" have every methods that we use in this part. cf. file if you want to see it all.

You can display the list of every existing phamily in the database curently selected.

```

1 list_name = sea_inphinity\var.get\_list\_name\_pham(-1)
2 print(list\_name)

```

### 3.5.2 Phages selection

### 3.5.3 Data completion

## 4 Results & Analyse

### 4.1 First result

#### 4.1.1 Tree

#### 4.1.2 Hosts

### 4.2 Database

#### 4.2.1 SEA

#### 4.2.2 Phages list integration

## 5 Conclusion

### 5.1 Problems encountered

#### 5.1.1 Phamerator Installation

#### 5.1.2 SEA database

#### 5.1.3 PhamDB Limitation jobs

#### 5.1.4 PhageDB.org

### 5.2 Perspectives

#### 5.2.1 Database population

#### 5.2.2 Resultats validation

#### 5.2.3 Results by host

# Bibliography

- [1] Docker. Docker official website. <https://www.docker.com/>.
- [2] YokAi Que MDPhD, Prof. Carlos Peña PhD, and Grégory Resch PhD. In silico prediction of phagebacteria infection networks as a tool to implement personalized phage therapy.
- [3] Ncbi. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3916379/>.
- [4] Phamerator. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-395>.



## **7 Annexes**