



MASTER OF SCIENCE  
IN ENGINEERING

UNIVERSITY OF APPLIED SCIENCES WESTERN SWITZERLAND  
MSE - SOFTWARE ENGINEERING

---

## Projet d'approfondissement

---

*Authors:*

Déruaz Vincent

*Supervisors:*

Déruaz Vincent

**Hes·SO**

Haute Ecole Spécialisée  
de Suisse occidentale

Fachhochschule Westschweiz

University of Applied Sciences and Arts  
Western Switzerland

Lausanne, April 26, 2016

# Contents

|          |                              |          |
|----------|------------------------------|----------|
| <b>1</b> | <b>Mise en place</b>         | <b>2</b> |
| <b>2</b> | <b>Phamerator</b>            | <b>3</b> |
| <b>3</b> | <b>Database migration</b>    | <b>4</b> |
| 3.1      | Completion of data . . . . . | 4        |
| <b>4</b> | <b>Commands</b>              | <b>5</b> |

# 1 Mise en place

## 2 Phamerator

## 3 Database migration

Phamator generate and populate a table, "pham", who describe the classification of genes in phamily. A phamily is a phamily of phages.

the "gene" table allow us to find wich gene is part of which phage. From this table we'll use the translation field to make a alignement and distance sequence analysis.

First try is made with 113 phages, 12482 genes and 2771 phams.

### 3.1 Completion of data

In the "phage" table somme of the data are missing. In deed, the "HostStrain" field can be to None. We wanted to complete those data.

The "Accession" field, from gene table, allow us to get the phage id from GeneBank.

**TODO:** Script python ???

- read sql -> get ids
- get missing informations
- write them on db

To do so, we need to get them from ...

We also wanted to know the HostSpecies. So, we modify the database to add a new field, "HostSpecies".

**TODO:** Script Sql to modify db

## 4 Commands

```
SELECT * FROM pham WHERE 1 GROUP BY name
```