Introduction and Installation

This document describes how to install and use the newly updated kClust Phameration script for creation of databases. The new script is available along with several other rewritten helper scripts from a bitbucket repository at:

https://bitbucket.org/chazbot/k_phamerate

Other Required Programs

The updated script requires three programs to be installed to function properly:

- **kClust -** ftp://toolkit.lmb.uni-muenchen.de/pub/kClust/
  Install kClust by downloading "kClust" from the above link. Simply place this executable in a folder and link it to a bin for global execution. It is highly recommended that you read the kClust paper to understand the new process.
- **kAlign2 -** http://msa.sbc.su.se/downloads/kalign/current.tar.gz
  Compile and install kAlign2 based on the README file in the distribution. Following these instructions creates the links necessary to execute the program globally.
- **HH-suite -** ftp://toolkit.genzentrum.lmu.de/pub/HH-suite/
  Download and extract the HH-suite tools to a directory, then link them into a bin for global execution. Alternatively you can add the "bin" directory in the distribution to your PATH. In order for global execution of hhconsensus and hhmake to work, you also must add the HHLIB variable to the PATH of the user in question. You can accomplish both of these requirements by appending the following to your user's ~/.profile

> **export HHLIB=/path/to/hhsuite/base/directory/**
> **PATH="$PATH:/path/to/hhsuite/bin/folder/"**

These programs should be installed and globally executable via the bin via links or PATH modifications. You should be able to type the commands to launch them – kClust, kAlign, hhmake and hhconsensus – globally and be greeted by usage instructions and/or version information.

Usage of Script

This new script only effects phameration of genes and phage deletion.

- You need to perform some changes to the SQL schema of a database for use with the new script, as phage deletion is now a smoother process. Refer to the file **k_schema_updates** for the relevant queries. Upon execution, you will be able to remove phages from the database using direct SQL query or an SQL editor, such as MySQLWorkbench:

  > **REMOVE FROM `phage` WHERE `phageid` = 'string';**

- As the Phameration no longer requires alignment information, you can shrink your databases substantially by truncating the scores_summary table. (see **k_schema_updates** for details)
- Once all of the above prerequisites are satisfied, back up your current database, and you can Phamerate it by executing the **k_phamerate.py** script with the following syntax:

  > **./k_phamerate.py DATABASE**

  The script runs an iteration of kClust, does alignments and extractions, then inserts new phamily designations after a second kClust iterations. Phamily names will be conserved if they exist in the database beforehand. **DO NOT INTERRUPT THE PROCESS**. If a Phameration is interrupted, restore from backup and try again. Contact us if you experience any issues!

- k_phamerate currently has issues setting colors with some orphams or phamilies created from orphams. Included in this repository is a script that will fix these errors for you. After every run of k_phamerate.py, you should run the following command:

  > **./phix_hue.py DATABASE**