Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Carthage

École Supérieure de la Statistique et de l'Analyse de l'Information

Institut Pasteur de Tunis

Rapport de Stage

d'Insertion à la Vie Professionnelle

Réalisé par:	**Encadré par:**	**Durée:**
Hannachi Samar (ESSAIT)	Souiai Oussema (IPT)	Du 31 Mai 2017 au 30 Juin 2017

Année Universitaire : 2017-2018

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué de prés ou de loin au succés de ce premier stage effectué.

D'abord, j'adresse mes remerciements à mon professeur, Mme Abdeljaoued Tej Inès, qui m'a beaucoup aidé lors de ma recherche de stage et qui a fait parvenir ma candidature à ce poste d'assistante de recherche et qui m'a fortement recommandée auprès des responsables.

Je remercie vivement mon maître de stage, Mr. Souiai Oussama, maître assistant à l'Institut Pasteur de Tunis et BioInformaticien pour son accueil, le partage de savoir et le temps passé ensemble mais également pour sa confiance et les différentes initiatives qu'il m'a permis d'entreprendre en me rattrappant si jamais je m'égarais de mon chemin. Il fut d'une aide précieuse dans les moments les plus frustrants en m'incitant toujours à repousser mes limites et à voir le monde à travers une nouvelle perspective.

Je tiens enfin à remercier toute l'équipe de recherche pour son accueil en tenant à m'offrir les meilleures conditions de travail possibles et pour leurs explications et éclairsissements.

Table des matières

Table des figures

Introduction

"Il y a des domaines où il n'y aura jamais de progrès : l'homme sera toujours mortel, il sera toujours soumis à la maladie."

Paul Virilio

Dans le cadre de mon stage d'insertion à la vie professionnelle, j'ai choisi d'effectuer mon stage dans un endroit qui me permetterait de contribuer à faire des avancées dans le domaine médicale et à appliquer mes outils et mon savoir-faire afin d'identifier les causes de maladies inconnues à ce jour et donc de participer à la réduction des maladies sur terre et cela tout en titillant ma curiosité et ma soif de savoir.

Je voue un grand intérêt aux initiatives dont l'objectif est de tirer de l'or à partir d'informations brutes et de transformer l'incompréhensible en compréhensible pour pouvoir en tirer une source bénéfique à l'humanité.

L'Institut Pasteur de Tunis (IPT), dont l'historique en dit long, m'a permis de travailler sur un projet qui vise à trouver les causes d'une maladie chez une certaine espèce animale et m'a permis de côtoyer différentes personnes spécialisées dans différents domaines.

Je décrirai en premier lieu l'institut et plus particulièrement le laboratoire dans lequel j'ai travaillé et ensuite les missions dont j'ai eu la responsabilité.

1 Présentation de l'Institut Pasteur de Tunis (IPT)

1.1 Historique de l'IPT

L'IPT est un centre de vaccination et de recherche scientifique au service de la santé de l'Homme placé sous la tutelle du ministère de la santé où il assume les missions fixées par la loi 58-35 portant son statut qui définit l'IPT comme "Établissement de Recherche Scientifique d'après les méthodes Pasteuriennes" ce qui le dote qu'un conseil d'administration et d'un conseil scientifique.
L'installation du centre de vaccination fut créé par décret Beylical en 1893 et construit en 1902 puis en 1906 la fondation des archives de l'IPT.
À partir de 1988, l'institut a recentré ses activités autour de sa vocation naturelle : la recherche scientifique qui grâce à un financement généreux de l'État s'est vu attribuer des cadres scientifiques, des équipes de recherche renforcées et des ressources technologiques.
Le travail de recherche s'est fortement concentré sur la biotechnologie pour qu'en fin 2007, l'IPT se voit identifié comme structure porteuse de projets dédiée aux biotechnologies appliquées à la santé [1].

L'IPT est en accord de coopération scientifique avec de nombreuses institutions scientifiques locales et étrangères et est membre du réseau internationale des instituts Pasteurs.

1.2 Missions de l'IPT

L'IPT a pour rôles :

*Effectuer toutes les enquêtes, missions, analyses ou recherches scientifiques intéressantes soit sur la santé publique humaine ou animale, soit sur le développement économique de la Tunisie.

*Assurer la synergie entre la recherche structurante et innovante et les applications industrielles pour le bien des populations.

1.3 Le Laboratoire de BioInformatiques, BioMathématiques et BioStatistiques (BIMS)

BIMS est un des neuf laboratoires de recherche de l'IPT dirigé par la biologiste «Alia Ben Kahla».

Positionnement scientifique

Les principaux axes de recherche scientifique de ce laboratoire tournent autour d'un souci d'intégration de l'ensemble des niveaux d'organisation du vivant avec un ancrage dans le domaine de la santé et cela grâce à une modélisation qui permet de formaliser un ensemble d'interactions intégratives aidant par la suite à appréhender les mécanismes à l'origine des phénomènes biologiques observés.

Problématique de recherche-développement

Le programme de recherche du laboratoire est composé de cinq projets allant de l'échelle moléculaire à l'échelle démographique en passant par l'échelle cellulaire et l'échelle individuelle.

Ces projets ont pour objectifs de :

*Développer des protocoles permettant la transformation des données biologiques brutes disponibles en information compréhensible.

*Développer des systèmes d'aide à la décision pour une meilleure gestion et contrôle des problèmes biologiques abordés.

*Recourir à des approches diverses qui permettent de fournir une multitude de méthodes d'abstraction qui permettent une meilleure compréhension des phénomènes biologiques et de trouver un dénominateur commun entre les problématiques et les hypothèses biologiques que la modélisation se charge de formaliser.

Projets du laboratoire

*Projet1 : Génomique fonctionnelle : Développement de protocoles bioinformatiques permettant l'identification de gènes candidats associés à un phénotype particulier et l'identification est modélisation des processus biologiques et voies métaboliques impliqués.

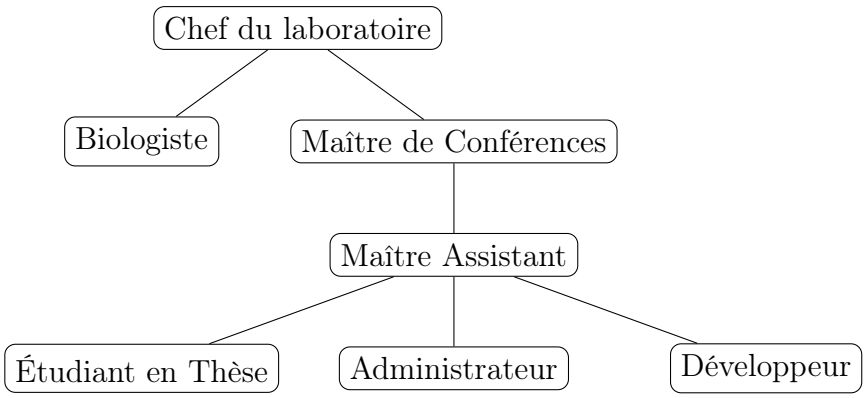
*Projet2 : Analyses statistiques de polymorphismes chimiques et de tendances métaboliques.

*Projet3 : Développement d'une plateforme informatique permettant d'intégrer et de traiter l'information biomédicale et de modéliser des profils de patients.

*Projet4 : Mise en place d'une plateforme d'aide à la décision pour le choix des traitements du cancer de la vessie : BLACkouT (BLAdder Cancer Therapy).

*Projet5 : Eco-évolution mathématique : évolution de traits d'histoire de vie épidémiologie sociale et vectorielle (obésité).

Composition du laboratoire



Protocole

*Les horaires pour les stagiaires à l'IPT vont de 9h du matin jusqu'à 14h de l'après-midi du Lundi au Vendredi tandis que tous les autres membres terminent à 17h.

*Aucune tenue n'est exigée et est préférable tout ce qui est décontracté.

2 Présentation du projet

Le projet s'intitulant "Les caractéristiques génomiques des souches virulentes de streptococcus pneumoniae de sérotype 1 [1] en nouvelle Caledonie" vise à étudier en profondeur l'épidémologie génétique de cette souche virulente et dans les meilleurs des cas à :

*** Évaluer la distribution des facteurs de virulence au sein de la population.**

*** Évaluer la distribution des phages au sein de la population.**

Les données mises à notre disposition sont sous forme de séquences génomiques de 73 extraits de streptococcus pneumoniae serotype 1 prélevées avant l'introduction du vaccin PCV13.

```
>4525_7#12_00001 transposase, ISSmi4
ATGATGATTCAACTAAGTGATTTAGGTCAAGTTCACATTGTTTGTGGCAAGACAGATATG
AGGCAGGGAATAGACTCATTAGCCTATGTAGTTAAAACCCACTTTGAATTGGATTCTTTC
TCCGGTCAAGTCTTTCTCTTTTGTGGTGGACGTAAAGACCGCTTTAAAGTCCTTTACTGG
GATGGTCAAGGATTTTGGCTACTATATAAACGCTTTGAGAACGGAAAACTGACTTGGCTA
AGTACAGAAAAGGATGTCAAAGCTCTCGCACCTGAACAAGTAGACTAG
>4525_7#12_00002 elongation factor Ts
ATGGCAGAAATTACAGCTAAACTTGTAAAAGAGTTGCGTGAAAAATCTGGTGCCGGTGTT
ATGGACGCTAAAAAAGCGCTTGTAGAAACAGACGGTGACATCGAAAAAGCGATTGAATTG
CTTCGTGAAAAAGGTATGGCTAAGGCAGCTAAGAAAGCTGACCGTGTTGCTGCAGAAGGT
TTGACTGGTGTTTATGTTAACGGTAATGTTGCAGCAGTTATTGAAGTAAACGCTGAAACT
GACTTCGTTGCAAAAAACGCTCAATTCGTTGAATTGGTAAATACTACAGCTAAAGTCATT
GCTGAAGGAAAACCTGCTAATAATGAAGAAGCTCTTGCTTTGATAATGCCTTCAGGTGAA
ACTCTTGAAGCTGCATACGTATCTGCAACAGCAACTATCGGAGAGAAAATCTCATTCCGT
CGCTTTGCATTGATTGAAAAAACAGACGCACAACACTTTGGAGCATACCAACATAACGGT
GGACGTATCGGTGTTATTTCAGTTGTTGAAGGTGGAGACGAAGCACTTGCTAAACAATTG
TCAATGCACATCGCAGCGATGAAACCAACAGTTCTTTCTTACAAAGAATTGGATGAGCAA
TTCGTTAAAGATGAGTTGGCACAATTGAATCACGTTATCGACCAAGACAACGAAAGCCGT
GCAATGGTTAATAAACCAGCTCTTCCACACTTGAAGTATGGATCAAAAGCTCAATTAACT
GATGATGTTATTGCTCAAGCTGAAGCTGACATCAAAGCTGAATTGGCTGCAGAAGCAAA
CCAGAAAAAATCTGGGACAAAATTATCCCAGGTAAAATGGATCGCTTCATGCTTGATAAT
ACTAAAGTTGACCAAGCTTACACACTTCTTGCACAAGTTTACATCATGGATGACAGCAAG
ACAGTTGAAGCATACCTTGAATCAGTTAACGCTTCGGTAGTTGAGTTTGCTCGCTTTGAA
GTTGGTGAAGGTATCGAGAAAGCTGCAAACGACTTCGAAGCTGAAGTTGCAGCTACAATG
GCAGCAGCCTTGAATAACTAA
>4525_7#12_00003 30S ribosomal protein S2
ATGGCAGTAATTTCAATGAAACAACTTCTTGAGGCTGGTGTACACTTTGGTCACCAAACT
CGTCGCTGGAATCCTAAGATGGCTAAGTACATCTTTACTGAACGTAACGGAATCCACGTT
ATCGACTTGCAACAAACTGTAAAATACGCTGACCAAGCATACGACTTCATGCGTGATGCA
GCAGCTAACGATGCAGTTGTATTGTTCGTTGGTACTAAGAAACAAGCAGCTGATGCAGTT
GCTGAAGAAGCAGTACGTTCAGGTCAATACTTCATCAACCACCGTTGGTTGGGTGGAACT
```

FIGURE 1 – Aperçu d'une séquence génomique

Ce projet est une collaboration entre l'Institut Pasteur de Tunis et le laboratoire de génétique au Malawi.

1. Une espèce de bactérie du genre Streptococcus. C'est un important agent pathogène chez l'Homme.

3 Étude Bibliographique

À défaut d'un certain bagage en biologie, indispensable à ce projet, j'ai passé la première partie de mon stage à effectuer un fastidieux effort bibliographique.

J'ai raflé les articles scientifiques qui avaient un lien de prés ou de loin avec mon projet, à tirer des similitudes avec ce qui avait déjà été effectué, à demander des éclaircissements aux collègues biologistes qui n'hésitaient pas à apporter leur aide, à me remettre sur le droit chemin si je m'en écartais et à me construire un certain savoir basique en biologie.

Une fois l'essentiel, qui m'a permis de tirer une véritable compréhension de ce qui était demandé acquis, je me suis tournée vers la maîtrise des logiciels exigés pour effectuer ce travail.

Le premier logiciel est un logiciel de détection de phages, PhiSpy[2] , dont l'installation et la maîtrise ont nécessité beaucoup de temps et de recherche.

Le deuxième utilisé est le logiciel R, une proposition personnelle de ma part, et qui n'a pas exigé une documentation vue que ma formation à l'ESSAI pendant la première année m'y initie.

Cette première étape s'est soldée par une petite évaluation sous la forme d'une présentation qui a défini l'avancement et les étapes à suivre.

4 Préparation de la base de données

Les données dispersées collectées par les biologistes étant non exploitables sous leur forme brute, j'ai eu pour première mission d'en tirer une base de données qui pourrait nous servir à tirer des résultats qui répondraient aux problématiques. Certains logiciels dédiés à la biologie telle que PhiSpy arrivent à détecter à partir des séquences d'acide nucléique les phages qui peuvent y être présents.

Après avoir maîtrisé le logiciel dit et l'avoir testé avec succès sur des données publiques j'ai dû y renoncer car il manquait à mon échantillon de données un code de reconnaissance essentiel à PhiSpy pour accomplir sa tâche.

Comme intermédiaire, j'ai eu recours à un serveur externe, Phast[3], auquel j'avais envoyé mes 73 échantillons un à un et qui me renvoyait les phages détectés dans chacun d'eux.

FIGURE 2 – Aperçu de phages detectés

Il fallait ensuite tirer de chaque résultat le code spécifique à chaque phage et faire coïncider les résultats. C'est-à-dire affecter des un quand le phage est présent pour une séquence d'ADN et zéro en cas d'absence.

Pour finir, j'ai éliminé les colonnes redondantes sans perdre de l'information en chemin grâce à un script R.

DNA\\PHAGES	NC_031941	NC_012756	NC_004996	NC_027396	NC_002666	NC_009737	NC_008376	NC_004584	NC_024357	NC_009
4525_7#12	1	1	1	1	1	1	1	1	1	
4564_2#8	0	1	1	1	1	1	1	1	1	
6308_7#14	0	1	1	1	1	1	1	1	1	
6308_7#15	0	1	1	1	1	1	1	1	1	
6308_7#16	0	1	1	1	1	1	1	1	1	
6308_7#17	0	1	1	1	1	1	1	1	1	
6308_7#18	1	1	1	1	1	1	1	1	1	
6308_7#19	0	1	1	1	1	1	1	1	1	
6308_7#20	0	1	1	1	1	1	1	1	1	
6308_7#21	0	1	1	1	1	1	1	1	1	
6309_7#10	1	1	1	1	1	1	1	1	1	
6309_7#12	1	1	1	1	1	1	1	1	1	
6570_4#3	1	1	1	1	1	1	1	1	1	
6570_4#4	1	1	1	1	1	1	1	1	1	
6570_4#5	0	1	1	1	1	1	1	1	1	
6570_4#9	1	1	1	1	1	1	1	1	1	
6570_4#11	0	1	1	1	1	1	1	1	1	
6570_4#13	1	1	1	1	1	1	1	1	1	
6570_4#15	0	1	1	1	1	1	1	1	1	
6570_4#17	0	1	1	1	1	1	1	1	1	
6570_4#18	1	1	1	1	1	1	1	1	1	
6851_2#76	1	1	1	1	1	1	1	1	1	
6851_2#81	1	1	1	1	1	1	1	1	1	
6851_2#85	1	1	1	1	1	1	1	1	1	
6851_2#87	0	1	1	1	1	1	1	1	1	
7712_7#49	1	1	1	1	1	1	1	1	1	

FIGURE 3 – Aperçu de la base de données finale

5 Classification ascendante hiérarchique

5.1 Définition

CAH : C'est une technique statistique visant à partitionner une population en différentes classes ou sous-groupes.On cherche à ce que les individus au sein d'une même classe soient le plus semblables possibles tandis que les classes soient le plus dissemblables possibles[4] .

5.2 Dendrogramme

La CAH va rassembler les observations de mon échantillon selon le critére de phages repérés pour produire un dendrogramme.

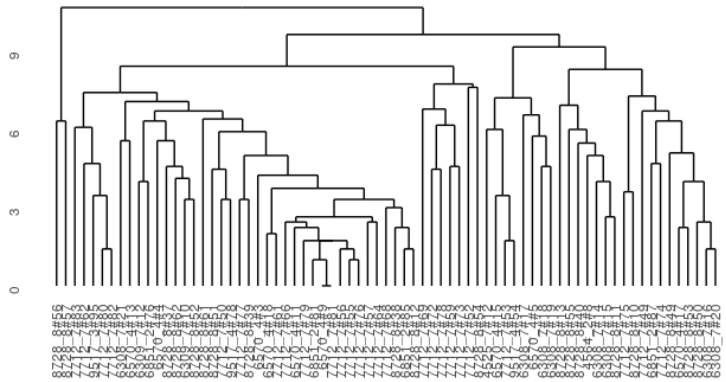


FIGURE 4 – Dendrogramme correspondant aux phages

Interprétation : Cette classification permet au biologiste de distinguer les caractéristiques communes aux différents échantillons et donc d'identifier potentiellement les caractéristiques génomiques des souches virulentes.

Conclusion

Ce premier stage a réussi à combler mes attentes aussi bien envers moi-même mais également envers l'organisme d'accueil. Ça m'a permis de me plonger dans le monde de la recherche de l'être vivant et de construire à partir de données encombrantes et inutilisables de l'information pure et dure.

Ce stage m'a également permis de comprendre la difficulté qui existe à passer d'un domaine à un autre et spécialement pour le monde de la biologie qui a exigé de ma part un effort de compréhension immense afin de pouvoir y appliquer mon savoir-faire.
Bien que je n'ai pu mener ce projet à terme, je suis fière d'avoir fait partie de son commencement et d'avoir pu aboutir à certains résultats qui serviront aux personnes qui suiveront.

Fortement enrichie de cette expérience, je souhaiterai, dans un prochain stage, m'orienter vers vers le secteur du "Machine Learning", un domaine en pleine expansion et dont le possibilités de réalisation sont gigantesques.

Bibliographie

[1] http ://www.pasteur.tn/.

[2] https ://www.ncbi.nlm.nih.gov/pmc/articles/PMC3439882/.

[3] http ://phast.wishartlab.com/index.html.

[4] http ://larmarange.github.io/analyse-R/classification-ascendante-hierarchique.html.