Universidad Autónoma de Madrid
Escuela politécnica superior

Máster en ...

Trabajo Fin de Máster

TÍTULO DEL TFM

Autor: Nombre Apellido1 Apellido2
Tutor: Nombre Apellido1 Apellido2
Ponente: Nombre Apellido1 Apellido2

MES 20xx

TÍTULO DEL TFM

Autor: Nombre Apellido1 Apellido2
Tutor: Nombre Apellido1 Apellido2
Ponente: Nombre Apellido1 Apellido2

Grupo de la EPS (opcional)
Dpto. de XXXXX
Escuela Politica Superior
Universidad Auta de Madrid
MES 20xx

Resumen

Resumen

Palabras Clave

Abstract

Key words

Agradecimientos

Índice general

Índice de Figuras

Índice de Tablas

1

Introducci

1.1. Motivacil proyecto

1.1. MOTIVACIL PROYECTO

Campylobacter jejuni es una bacteria Gram negativa que, a pesar de tener unas condiciones complicadas de crecimiento [1], es la zoonosis bacteriana que produce un mayor nmero de intoxicaciones alimentarias en los pas tanto desarrollados como en v de desarrollo. Por ejemplo, en la EU en el a16 se declararon del orden de 250.000 casos comprobados [2]. El coste debido a la campilobacteriosis se estima en la EU en torno a 2,4 billones de euros anuales. La fuente de contaminacis habitual es el consumo de carne de pollo poco cocinada [3]. El grupo de investigacicnofood lleva varios anvestigando sobre las fuentes de contaminaci este microorganismo a lo largo de la cadena alimentaria [1] [3] [4]. En la actualidad se dispone de una colecci Campylobacter spp. de alrededor de 2000 cepas. Con el fin de obtener una informacis precisa sobre la persistencia de este microorganismo a lo largo de la cadena alimentaria, se han aislado varios genotipos persistentes en el matadero. De estos se han secuenciado con un equipo MiSeq (Illumina) 45 de ellas.

El proyecto consiste en disen workflow que permita, a partir de los datos obtenidos en formato fastq proporcionados por el equipo, conseguir realizar las fases de trimming y evaluaci la calidad de las secuencias obtenidas, obtenci contigs, assembling y anotaciara poder tener la informaci la secuencia de genes del genoma completo [5] [6] [7] de las cepas de Campylobacter secuenciadas. En la actualidad existen varios programas desarrollados por varios grupos de investigaciternacional que realizan las funciones demandadas. Se trata de buscar la solucis eficaz y fl de implementar y que ds mejores resultados, por lo que habre comparar diferentes programas y estrategias. Adicionalmente, se requiere incorporar en este workflow o en ansis paralelos [8], la posibilidad de detectar insertos de origen viral y/o plidos en el genoma y herramientas que permitan la comparacipida de los genomas de las distintas cepas aisladas, algunas de ellas pertenecientes a cepas altamente clonales. Esta herramienta se ha demandado por parte de un grupo sin conocimientos informcos, por lo que se requiere desarrollar un entorno de fl uso por su parte.

El proyecto plantea una colaboracitre los grupos ADMIRABLE y TECNOFOOD de la Universidad de Burgos. Especializados en informca y ciencia y tecnologe los alimentos respectivamente. Dada esta combinaci disciplinas, el proyecto se encuentra en el marco de los trabajos

considerados dentro del campo de la bioinformca.

1.2. Objetivos y enfoque

1.2. OBJETIVOS Y ENFOQUE

El objetivo principal del proyecto es el desarrollo del workflow que permita el ansis de las cepas de Campylobacter jejuni, para lo que se utilizaralaxy y las herramientas disponibles en la "tool shed"(conjunto de herramientas que ofrece Galaxy para su instalaciara cada paso. Galaxy es una herramienta que permite ansis computacionales de datos biolos. El segundo objetivo se centra en crear una herramienta grca a trave la cual facilitar el uso de este workflow, que se llevarabo utilizando Python, Qt y la API de Galaxy. Adem para facilitar el despliegue y ejecuci la herramienta desarrollada, se va a crear un contenedor Docker. Por lo tanto, el siguiente objetivo parcial sersarrollar la imagen Docker que sirva de base. Finalmente, se desea tener alguna forma sencilla de tratar los datos de salida del workflow, por lo que dentro de la interfaz grca se incluirerramientas con las que gestionar toda la informacisultante en forma de grcos, tablas tipo hoja de culo, formatos pdf, etc.

Objetivos del proyecto:

- Desarrollar el workflow necesario para el ansis de las cepas en Galaxy
- Crear un sistema Docker sobre el que desplegar el proyecto
- Desarrollar una interfaz grca para simplificar la utilizaci la aplicacir un sistema de gestincilla de los datos de salida.

1.3. Metodolog plan de trabajo

1.3. METODOLOG PLAN DE TRABAJO

1.3.1. Metodolog

La metodologtilizada en el desarrollo del proyecto, dada su cercann su estructura a un proyecto de software tradicional, ser tipo l. Se basar el tipo de metodologcrum, con reuniones en cada sprint. La carga de trabajo, como aproximacialta de conocer ciertos requisitos que puedan surgir durante el desarrollo, se dividir la estructura siguiente.

1.3.2. Plan de Trabajo

Sprint 1 (18/9/2018 - 3/10/2018)

El primer sprint ha estado centrado tanto en definir con mxactitud la direccil proyecto como en un primer acercamiento a las principales herramientas con las que va a desarrollarse.

Tras unos primeros pasos con Galaxy [9] y Docker [10], se ha tomado como referencia el trabajo Bioinfworkflow de Sergio Chico [11] como base para la imagen Docker del proyecto. Dado que el proyecto de Github daba algunos problemas en la instalacie ha desarrollado un script propio que produce los mismos resultados.

Una vez se ha tenido disponible la imagen de Docker, el sprint se ha centrado en algunos aspectos importantes para partes futuras del desarrollo. Entre ellos destaca la investigacierca

Tareas / subtareas	Horas
T1. Desarrollo de la imagen Docker	80
T1.1 Adaptar la imagen previa orientada a Galaxy	50
T1.2 Permitir desplegar la imagen en un servidor	30
T2. Desarrollo del workflow en Galaxy	110
T2.1. Selecci las herramientas	50
T2.2. Configuraciramica	30
T2.3. Habilitar ejecucil workflow de manera programca	30
T3. Desarrollo de la interfaz grca	60
T4. Desarrollo del sistema de tratamiento de datos de salida	50
TOTAL HORAS	300

Cuadro 1.1: Plan de Trabajo

del formato de los workflows de Galaxy (.ga) ya que en un futuro sercesario generar este tipo de ficheros para introducirlos en Galaxy. Tambiesulta relevante la investigacierca de las posibilidades que ofrece la API de Galaxy [11] y su utilidad en Bioblend [12], que nos facilitan la opci utilizar Galaxy sin necesidad de hacerlo a trave su interfaz.

Sprint 2 (4/10/2018 - 17/10/2018)

La primera semana de este sprint ha estado dirigida a lograr una imagen Docker de Galaxy que contenga un set de herramientas bcas para formar un primer workflow. Se han valorado varias opciones de instalaci las que se han utilizado tanto la imagen bca de Galaxy [13] como la imagen de Bioinfworkflow [11]. Finalmente se ha optado por utilizar Bioinforworkflow ya que parte de las herramientas necesarias ya estaban incluidas. Para realizar esta tarea se ha creado un nuevo fichero Dockerfile asmo el listado de herramientas necesarias para su instalaci

Sprint 3 (18/10/2018 - 31/10/2018)

El sprint ha estado centrado en la correcta ejecucil workflow con las herramientas iniciales desde Galaxy. Durante el proceso de configuracin surgido varias complicaciones que han impedido terminar el workflow completo en este sprint. En un principio han surgido problemas con el filtrado de calidad utilizando Prinseq. Este problema no ha llegado a ser resuelto en este sprint a falta de tratar el tema con el grupo de Tecnofood. A continuaci encontraron ciertos problemas con el formato de salida de la herramienta Prokka. A pesar de que la salida estrcada como formato gff3, un partro interno lo etiquetaba como gff. Esto impedue la salida de Prokka fuese introducida como input en las herramientas siguientes.

Dados estos errores, se decidibajar en paralelo con la API de Galaxy desde Python para intentar ejecutar tanto las herramientas como el workflow de una manera menos restringida. Finalmente se ha llevado a cabo el desarrollo necesario para subir los ficheros a un historial y ejecutar cada una de las herramientas del workflow desde Python.

Sprint 4 (1/11/2018 - 14/11/2018)

La prioridad en este punto se ha centrado en completar el workflow desde Galaxy. Han surgido varios problemas en esta tarea. La primera es un bug en Mac por el cual los archivos eliminados dentro de Docker no se eliminan del todo y quedan fijados en un fichero residual. Esto implica que cada cierto tiempo hay que eliminar la imagen completa de Docker para poder liberar espacio, dado el gran tama los archivos con los que se trabaja. Debido a ello, la tarea de completar el workflow se ha visto retrasada. Adem la ejecuci la herramienta Roary a trave Galaxy ejecuta sin errores pero no devuelve ningn resultado, lo que ha impedido continuar con la parte final del workflow.

Ademe la tarea ya comentada, en este sprint se ha trabajado en el acceso al contenedor Docker desde otro ordenador en una red local. Con el objetivo de desplegar el servicio en un servidor.

Tambie ha tomado contacto con la interfaz grca, realizando unas pruebas en las que simplemente se muestra alguna informacitra de la API de Galaxy en etiquetas creadas con PyQt.

Sprint 5 (15/11/2018 - 28/11/2018)

Al igual que en los sprints anteriores, gran parte de la carga de trabajo se ha centrado en resolver problemas en la ejecucil workflow a trave Galaxy. Se han realizado numerosas ejecuciones para comprobar si las salidas de cada herramienta eran las correctas. Esto ha servido para concluir que, al parecer, un fallo en la herramienta Roary incluida en Galaxy, impide que los ficheros retornados tengan contenido. Esto se ha comprobado a trave la ejecuci Roary standalone con los mismos datos de entrada y los mismos partros, obteniendo de esta manera los ficheros correctos.

Para ahorrar en tiempos de ejecuci han utilizado los ficheros fasta ya generados previamente, no los generados con la herramienta Spades de nuestro propio workflow. En el pro sprint se tratar integrar la parte previa a los pasos que ya son correctos.

Tambie ha ado un fichero .gitignore para evitar la existencia de ficheros irrelevantes en el repositorio.

Parte del trabajo de este sprint se ha destinado a la redacci la propuesta de proyecto y de la introduccia documentacil mismo.

Sprint 6 (29/11/2018 - 12/12/2018)

El trabajo de este sprint se ha centrado en fragmentar el proceso del workflow lo mmo posible para detectar d se estroduciendo los errores. Inicialmente se ha acortado el workflow hasta el paso de ensamblaje con Spades, ya que los problemas surg en este punto. Posteriormente se ha ejecutado el workflow individualmente en lugar de por colecciones. De esta manera, se ha detectado que el problema se esta dando en la ejecuci Spades con dos secuencias concretas: 590 y 443. Tras investigar probando varias ejecuciones con diferentes partros, se ha llegado a la conclusi que no era un fallo de configuraci la herramienta, sino de hardware. Al ceder a docker una cantidad mayor de memoria RAM (8 Gb), el problema se ha solucionado. A continuacie ha pasado a ejecutar de nuevo por colecciones hasta el paso de Spades, para comprobar si con este cambio ha sido suficiente para que la ejecucia correcta con este formato.

Tambie ha realizado una modificacil fichero .gitignore para mantener los archivos .pdf generados por LaTeX.

Sprint 7 (13/12/2018 - 26/12/2018)

La tarea principal de este sprint ha sido la creaci una herramienta Roary que poder integrar en Galaxy. Se valorpci crear una nueva imagen Galaxy instalando la herramienta desde la creaciicial de Docker, pero finalmente se ha optado por subir esta versi Roary al tool shed de Galaxy.

A continuacie han realizado varias comprobaciones del funcionamiento de la integraci esta herramienta, con resultados positivos. Sin embargo, al ejecutar el workflow completo, parecen surgir de nuevo errores en la parte de Prokka, probablemente provocados por la ejecuci Spades.

2

Reconocimiento de iris. Estado del arte

2.1. Introducci

2.2. Historia, nacimiento y evoluci

2.2. HISTORIA, NACIMIENTO Y EVOLUCI

2.3. La anatomel ojo

2.3. LA ANATOMEL OJO

2.3.1. Aspectos diferenciadores del iris

2.4. Adquisicil Iris

2.4. ADQUISICIL IRIS

2.4.1. Introducci

2.4.2. Esquemas de adquisiciadicionales

2.4.3. Consideraciones sobre la iluminaci

2.4.4. Posicionamiento del Iris

2.4.5. Sistemas comerciales de adquisici

2.5. Localizaciegmentacil Iris

2.5. LOCALIZACIEGMENTACIL IRIS

2.5.1. Introducci

2.5.2. Metodologe J. Daugman y derivadas

2.5.3. Metodologe R. Wildes y derivadas

2.5.4. Otras metodolog

2.5.5. Comparativa de metodolog

2.5.6. Detecci pesta ruido

2.6. Normalizacil tama

2.6.1. Daugman's Rubber Sheet Model

2.6.2. Image Registration

2.6.3. Normalizaci ulo

2.6.4. Mejora del contraste y eliminaci ruido

2.7. Algoritmos de Codificaci

2.7.1. Metodologe Daugman: Filtros de Gabor

2.7.2. Metodolog alternativas a la de Daugman

Filtros Log-Gabor

Wavelets

Haar Wavelet

Transformada Discreta del Coseno (DCT)

2.7.3. Metodolog de Wildes. Vectores de caractericas reales (no binarios)

2.8. Algoritmos de Matching

2.8.1. Introducci

2.8.2. Distancia de Hamming

2.8.3. Distancia eucla ponderada

2.8.4. Correlacirmalizada

2.9. Problemca y retos futuros

2.9.1. Segmentaci

2.9.2. Captura ideal no invasiva

2.10. Competiciones o Evaluaciones de Iris

2.10.1. The Iris Challenge Evaluation (ICE)

2.10.2. The Noisy Iris Challenge Evaluation (NICE)

2.11. Bases de datos

2.11.1. CASIA

2.11.2. BioSec Baseline y BioSecurID

3

Sistema, diseesarrollo

3.1. Segmentaci

3.2. Normalizaci

3.2. NORMALIZACI

3.3. Codificaci

3.3. CODIFICACI

3.4. Matching

3.4. MATCHING

4

Experimentos Realizados y Resultados

4.1. Bases de datos y protocolo

4.2. Sistemas de referencia

4.3. Escenarios de pruebas

4.4. Experimentos del sistema completo

5

Conclusiones y trabajo futuro

Glosario de acros

- **IS**: Iris Subject
- **DCT**: Discrete Cosine Transform
- **WED**: Weighted Euclidean Distance

Bibliografía

[1] Lourdes García-Sánchez, Beatriz Melero, Isabel Jaime, Marja-Liisa Hänninen, Mirko Rossi, and Jordi Rovira. Campylobacter jejuni survival in a poultry processing plant environment. *Food Microbiology*, 65:185–192, aug 2017.

[2] Campylobacteriosis - Annual Epidemiological Report 2016 [2014 data].

[3] Lourdes García-Sánchez, Beatriz Melero, Ana Ma Diez, Isabel Jaime, and Jordi Rovira. Characterization of Campylobacter species in Spanish retail from different fresh chicken products and their antimicrobial resistance. *Food Microbiology*, 76:457–465, dec 2018.

[4] Beatriz Melero, Pekka Juntunen, Marja-Liisa Hänninen, Isabel Jaime, and Jordi Rovira. Tracing Campylobacter jejuni strains along the poultry meat production chain from farm to retail by pulsed-field gel electrophoresis, and the antimicrobial resistance of isolates. *Food Microbiology*, 32(1):124–128, oct 2012.

[5] Clifford G. Clark, Chrystal Berry, Matthew Walker, Aaron Petkau, Dillon O. R. Barker, Cai Guan, Aleisha Reimer, and Eduardo N. Taboada. Genomic insights from whole genome sequencing of four clonal outbreak Campylobacter jejuni assessed within the global C. jejuni population. *BMC Genomics*, 17(1):990, dec 2016.

[6] Ann-Katrin Llarena, Eduardo Taboada, and Mirko Rossi. Whole-Genome Sequencing in Epidemiology of Campylobacter jejuni Infections. *Journal of clinical microbiology*, 55(5):1269–1275, may 2017.

[7] S. Zhao, G. H. Tyson, Y. Chen, C. Li, S. Mukherjee, S. Young, C. Lam, J. P. Folster, J. M. Whichard, and P. F. McDermott. Whole-Genome Sequencing Analysis Accurately Predicts Antimicrobial Resistance Phenotypes in Campylobacter spp. *Appl. Environ. Microbiol.*, 82(2):459–466, jan 2016.

[8] C. P. A. Skarp, O. Akinrinade, A. J. E. Nilsson, P. Ellström, S. Myllykangas, and H. Rautelin. Comparative genomics and genome biology of invasive Campylobacter jejuni. *Scientific Reports*, 5(1):17300, dec 2015.

[9] Galaxy. `https://usegalaxy.org/`.

[10] Docker - Build, Ship, and Run Any App, Anywhere. `https://www.docker.com/`.

[11] Sergio Chico. :whale: Docker with Galaxy for Bioinformatic Bacterial Sequencing Workflows: Serux/docker-galaxy-BioInfWorkflow. `https://github.com/Serux/docker-galaxy-BioInfWorkflow`, June 2018. original-date: 2018-05-17T01:10:05Z.

[12] Galaxy API. `https://galaxyproject.org/develop/api/`.

[13] Björn Grüning. :whale::bar_chart::books: Docker Images tracking the stable Galaxy releases.: bgruening/docker-galaxy-stable. https://github.com/bgruening/docker-galaxy-stable, October 2018. original-date: 2014-08-12T13:26:14Z.

A

Manual de utilizaci

B

Manual del programador