Internship proposal - Master 2 Bioinformatics

Specialisation Software Development and Data Analysis

Internship period: Jan-Jun 2024

|  |  |
| --- | --- |
| **Internship title** | Genomic Characterization of North American Isolates of Streptococcus suis Serotype 1/2: A Bioinformatics Approach for Understanding Population Structure |
| **Internship supervisor(s)** | Nahuel Fittipaldi, MSC, PhD; Marcelo Gottschalk, DVM, PhD |
| **Laboratory acronym(s)** | Fittipaldi and Gottschalk laboratories |
| **Laboratory city(ies)** | St-Hyacinthe, Quebec, Canada |
| **Web site(s)** | <https://fmv.umontreal.ca/etudes/personnel-enseignant/fiche/in/in33239/sg/Nahuel%20Fittipaldi/> |
| **Team(s)** | Dr. Fittipaldi leads a team currently comprising a research assistant, one PhD and 2 MSc students focusing on the |
| **E-mail(s)** | n.fittipaldi@umontreal.ca |
| **Internship description** | **Project Overview:**  The primary goal of this project is to comprehensively characterize the population structure of Streptococcus suis serotype 1/2, utilizing a cohort of 115 genomes derived from North American isolates. Streptococcus suis is a pathogen of substantial economic and health concern, causing severe diseases in pigs and zoonotic infections in humans. A more detailed understanding of the genomic makeup and population structure of this bacterium can lead to improved diagnostic strategies, therapies, and preventive measures.  **Methods:**  The project will employ an array of bioinformatics approaches for analysis:   1. **Genome Assembly and Annotation:** As a preliminary step, each genome will be assembled using tools like SPAdes, and annotation will be performed using Prokka or a similar tool. 2. **Multi-Locus Sequence Typing (MLST):** To identify sequence types (STs), each genome will undergo MLST. The data obtained will be analyzed for its relevance to the pathogenicity and geographic distribution of the strains. 3. **Whole Genome SNP Analysis:** SNPs will be identified across the 115 genomes using tools like GATK or SAMtools. SNP data will be used to understand genetic variation and infer evolutionary relationships among the isolates. 4. **Phylogenomic Analysis:** Phylogenetic trees will be constructed using tools like RAxML or IQ-TREE based on whole-genome SNP data. This will help understand the evolutionary relationships between isolates. 5. **Pan-Genome Analysis:** To characterize the core and accessory genomes, Roary or a similar tool will be used. The pan-genome analysis will provide insights into gene gain/loss events, and the accessory genome will help understand the variability within the population. 6. **Antibiotic Resistance and Virulence Gene Profiling:** Tools like ResFinder and VirulenceFinder will be employed to identify the presence of antibiotic resistance and virulence genes. This will aid in determining potential risks associated with these isolates.   **Expected Outcomes:**  By the end of the project, a comprehensive characterization of the population structure of Streptococcus suis serotype 1/2 derived from North American isolates will be achieved. It will yield important insights about the pathogen's genetic diversity, evolution, virulence factors, and antibiotic resistance patterns. The knowledge generated could directly contribute to public health measures and potentially lead to better strategies for treatment and prevention of infections caused by this pathogen |
| **Expected skills** | **Three or more of the following:**   1. **Genomics:** A foundational understanding of genomics and molecular biology is desiredl. 2. **Bioinformatics:** Familiarity with various bioinformatics tools and databases, such as those used for genome assembly, annotation, and sequence analysis. 3. **Programming:** Proficiency in languages like Python, R, or Perl, to manipulate data and automate tasks. 4. **Statistical Analysis:** Skills to interpret complex biological data and perform statistical analysis. 5. **Data Visualization:** Ability to use tools for visualizing complex genomic data, such as ggplot2 in R or matplotlib in Python. 6. **Critical Thinking:** Ability to design experiments, troubleshoot, and interpret complex genomic data. 7. **Communication Skills:** Ability to clearly articulate findings both verbally and in written format. 8. **Teamwork:** Ability to collaborate effectively with a team of multidisciplinary scientists. |
| **Confidential (yes/no)** | No |
| **Informatics resources** | We are excited to share that our team is equipped with state-of-the-art hardware and cutting-edge bioinformatics software to enable efficient and comprehensive analysis of bacterial genomic data. Our existing resources combined with our team's expert knowledge in genomics, statistical analysis, and data visualization make us well-positioned to tackle complex bioinformatics challenges. This environment offers an unparalleled opportunity for trainees to advance their skills, work with advanced computational tools, and make meaningful contributions to real-world scientific investigations. We invite you to apply and be a part of our vibrant team that is at the forefront of bacterial genomics research. |
| **Potential PhD project (yes/no)** | Yes |
| **References (2 and 3)** | <https://scholar.google.com/citations?hl=en&user=fZmBhq4AAAAJ&view_op=list_works&sortby=pubdate>  <https://scholar.google.com/citations?hl=en&user=IKMDfo0AAAAJ&view_op=list_works&sortby=pubdate> |