

SYLLABUS

COURSE NUMBER:

COURSE TITLE:

Genomic surveillance of antimicrobial resistance

COURSE DESCRIPTION:

The emergence and spread of antimicrobial resistant pathogens threaten hospitals and communities at every level and are now a major global threat to public health. Whole genome sequencing (WGS) of antimicrobial resistant pathogens is becoming more and more relevant for the medical sector. This graduate course will provide students a basis to understand and be acquainted with applications of WGS with Illumina and Nanopore technologies in surveillance of bacterial pathogens for infection control and antibiotic stewardship. We will make use of the current leading-edge bioinformatic tools and databases for species-level pathogen identification, molecular typing methods, and antimicrobial resistance detection linking epidemiological data. We will use real dataset, vancomycin- and daptomycin-resistant *Enterococcus faecium* Illumina and Nanopore data generated with the support of UAMS Barton Pilot Grant. The class will meet once per week for a 50 minute which will be combined with lecture and hands-on computer lab session. Students will need access to a computer and open-source software required for the course will be installed during lectures.

PRE-REQUISITES:

BMIG 5003 Computational Methods for Informatics

GENERAL INFORMATION:

CREDITS:

1

SEMESTER:

Spring

MEETING TIME:

Mondays 9:00am-10:00am

LOCATION:

UAMS Main Campus – Face-to-Face

FACULTY:

Se-Ran Jun, PhD

Guest Lecturers

OFFICE: Biomed II, 363-2

OFFICE HOURS: By appointment only

PHONE: 501-686-6025

EMAIL: sjun@uams.edu

SPECIAL ASSISTANCE: Students who believe they may need accommodations in this class based on mental or physical impairments must contact the Associate Dean for Academic Affairs at (501) 686-5730 to schedule an appointment to discuss your needs. Please make arrangements as soon as possible so accommodations can be made in a timely manner.

COURSE OBJECTIVES: Upon successful completion of this course, the student is able to:

1. Describe how to do de novo assembly from raw reads to contigs using Illumina and Nanopore data
2. Apply tools for de novo assembly and hybrid de novo assembly with Illumina and Nanopore data
3. Identify and apply tools and reference database for species-level pathogen identification
4. Describe methods of molecular epidemiology tools (MLST, cgSNP, and cgMLST)
5. Utilize MLST, cgSNP, and cgMLST tools for molecular typing
6. Demonstrate how to construct minimum spanning tree based on SNPs
7. Demonstrate how to construct phylogenomic tree and explain the relatedness of isolates in combination
8. Use an online tool, interactive tree of life (iTOL) for the display, annotation and management of phylogenomic trees
9. Identify and apply tools and reference databases for antibiotic resistance detection
10. Give examples of the applications of whole genome sequencing with Illumina and Nanopore data to surveillance of antimicrobial resistant bacterial pathogens
11. Document analysis pipeline via for example, Jupyter Notebook.

MAJOR TOPICS: The course will be split into four sections. Each focusing on core concepts of genomic surveillance for infection control and antibiotic stewardship.

1. Part 1: De novo assembly with Illumina and Nanopore raw data. This part will explore cutting-edge tools for de novo assembly and hybrid de novo assembly with Illumina and Nanopore data and evaluate the results.
2. Part 2: Species-level pathogen identification. This part will explore a comprehensive reference genome database, and a microbial classification engine to identify pathogen at the species level.
3. Part 3: Typing methods and phylogenomic tree construction for infection control. This part will explore most popular molecular epidemiology tools for typing and build phylogenomic trees to elucidate relatedness of pathogen strains for infection control.
4. Part 4: Antibiotic resistance detection for antibiotic stewardship. This part will focus on how to perform in silico antibiotic resistance detection based on homology models and SNP models to detect genes and mutations known to be associated with antibiotic resistance.

COURSE OUTLINE:

Reports and Assignments: At the end of each of the four major parts, a summary of methods, tools and databases discussed in the lectures will be submitted by the student. Further, the student will submit documented scripts to accomplish course goals.

Course Project: At the end of the course the student will submit a written report of results obtained from all four major topics covered in the course. The student will be provided whole genome sequencing data from Illumina and Nanopore technologies if necessary.

EVALUATION:

This is a graded course. Grades will be assigned based on the course average according to the following scale: A (93-100), B (84-92), C (75-83), D (65-74), Fail (≤ 64). The course average will be comprised of course assignments, the course project, the annotated bibliography, and class participation.

Reports	30%
Assignments	25%
Course Project	30%
Class Participation	15%

STUDENT FEEDBACK / EVALUATION OF INSTRUCTION:

At the end of the semester, you will be asked to fill out an online “Course and Instructor Evaluation” survey. The survey is anonymous and the data is de-identified and aggregated by the Academic Program Director; instructors do not receive survey results until all grades are turned into the evaluation system so opinions and comments are expressed freely on the survey do not have an effect on student’s grades. Instructor(s) and course evaluation provide valuable feedback for the program and courses taught at the Department of Biomedical Informatics. In any time during the semester your feedback is extremely important and encouraged throughout the course and the survey.

ACADEMIC HONESTY:

Academic honesty is expected at all times. All graded work must be your own unless otherwise specified in the assignment. Fair credit must be given to others for their work on team assignments by including a statement of contributorship (see ICMJE guidelines for authorship). Academic dishonesty, such as, but not limited to cheating, plagiarism, using the work of others without permission, and acknowledgement and forgery will result in an automatic zero for the assignment and may result in a failing grade in the course, loss of graduate funding, and dismissal from your degree program.

ATTENDANCE:

Students are expected to be diligent in the pursuit of their studies and in their class attendance. Students have the responsibility of making arrangements satisfactory to the instructor regarding all absences. Such arrangements should be made prior to the absence, if possible. Policies of making up work missed as a result of absence are at the discretion of the instructor and students

should verify the policies of their instructors at the beginning of each semester. Attendance may affect the class participation grade.

ADDITIONAL RESOURCES:

Please refer to the Graduate School Student Handbook for resources at your disposal available around the campus. Students are expected to have access to a personal computer on which they can load and install the software, or access the web sites or online resources, used to accomplish the exercises. Course materials will be drawn from articles or online resources. Students with limited programming experience can optionally refer to the following text as reference.

Mac OS X UNIX Tutorial for Beginners

<https://faculty1.coloradocollege.edu/~sburns/UnixTutorial/index.html>

Linux Command Line Basics

<https://www.udacity.com/course/linux-command-line-basics--ud595>

Anaconda / Miniconda:

<https://www.anaconda.com>

<https://conda.io/miniconda.html>

Jupyter Notebook / Jupyter Lab

<https://jupyter.org/>

Biopython

<https://biopython.org/>

Python Data Science Handbook

<https://jakevdp.github.io/PythonDataScienceHandbook/>

Tentative Course Schedule – The class will meet once per week for a 50 minute which will be combined with lecture and hands-on computer lab session at the same time.

Week 1	<p>Topics: Course introduction and overview</p> <p>Learning Objectives:</p> <ul style="list-style-type: none"> ▪ Understand need for whole genome sequencing for hospital/community outbreak analysis ▪ Understand need for whole genome sequencing for antibiotic stewardship ▪ Get familiar with and install required tools: Unix/Miniconda/Jupyter Lab/R <p>Reading:</p> <ul style="list-style-type: none"> ▪ Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis by Quainoo et al, Clin Microbiol Rev. 2017;30:1015.
Week 2	<p>Topics: Next Generation Sequencing (NGS) technologies for Whole Genome Sequencing (WGS) and de novo assembly (I)</p> <p>Lectures Objectives:</p> <ul style="list-style-type: none"> ▪ Understand Illumina data in terms of read length, read depth and error rate ▪ Explore a pipeline for de novo genome assembly for Illumine data <p>Reading:</p> <ul style="list-style-type: none"> ▪ https://www.melbournebioinformatics.org.au/tutorials/tutorials/assembly/assembly-protocol/ ▪ The present and future of de novo whole-genome assembly by Sohn et al, Briefings in Bioinformatics. 2018;19:23.
Week 3	<p>Topics: Next Generation Sequencing (NGS) technologies for Whole Genome Sequencing (WGS) and de novo assembly (II)</p> <p>Learning Objectives:</p> <ul style="list-style-type: none"> ▪ Understand Nanopore data in terms of read length, read depth and error rate ▪ Explore a pipeline for de novo genome assembly for Nanopore data <p>Reading:</p> <ul style="list-style-type: none"> ▪ https://www.polarmicrobes.org/tutorial-nanopore-analysis-pipeline/

	<ul style="list-style-type: none"> ▪ https://galaxyproject.github.io/training-material/topics/assembly/tutorials/unicycler-assembly/tutorial.html ▪ https://readthedocs.org/projects/denbi-nanopore-training-course/downloads/pdf/latest/ ▪ The present and future of de novo whole-genome assembly by Sohn et al, Briefings in Bioinformatics. 2018;19:23.
Week 4	<p>Topics: Topics: Next Generation Sequencing (NGS) technologies for Whole Genome Sequencing (WGS) and de novo assembly (III)</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Explore hybrid de novo assembly of Illumina and Nanopore data <p>Reading:</p> <ul style="list-style-type: none"> ▪ https://github.com/rrwick/Unicycler ▪ The present and future of de novo whole-genome assembly by Sohn et al, Briefings in Bioinformatics. 2018;19:23. ▪ Nanopore sequencing reads improve assembly and gene annotation of the Parochlus steinenii genome by Shin et al, Scientific Reports, 2019;9:
Week 5	<p>Topics: Topics: Next Generation Sequencing (NGS) technologies for Whole Genome Sequencing (WGS) and de novo assembly (IV)</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Assembly evaluation with Icarus ▪ Compare de novo assemblies and hybrid de novo assembly <p>Reading:</p> <ul style="list-style-type: none"> ▪ Icarus: visualizer for de novo assembly evaluation Mikheenko et al, Bioinformatics, 2016;32:3321. ▪ Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing Goldstein et al, BMC Genomics, 2019;20:23. ▪ GMASS: a novel measure for genome assembly structural similarity Kwon et al, BMC Bioinformatics, 2019;20:147.
Week 6	Topics: Species-Level Pathogen Identification (I)

	<p>Lectures:</p> <ul style="list-style-type: none"> ▪ Reference genome database ▪ Integrate local genome database into reference genome database <p>Reading:</p> <ul style="list-style-type: none"> ▪ Centrifuge: rapid and sensitive classification of metagenomic sequences by Kim et al, Genome Res, 2016;26:1721. ▪ https://ccb.jhu.edu/software/centrifuge/manual.shtml
Week 7	<p>Topics: Species-Level Pathogen Identification (II)</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Centrifuge classifier <p>Reading:</p> <ul style="list-style-type: none"> ▪ Centrifuge: rapid and sensitive classification of metagenomic sequences by Kim et al, Genome Res, 2016;26:1721. ▪ https://ccb.jhu.edu/software/centrifuge/manual.shtml
Week 8	<p>Topics: Genome annotation</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Genome annotation: PROKKA <p>Reading:</p> <ul style="list-style-type: none"> ▪ Prokka: rapid prokaryotic genome annotation Seemann T, Bioinformatics, 2014;30:2068. ▪ https://github.com/tseemann/prokka
Week 9	<p>Topics: Molecular Epidemiology Tools: MLST</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Multilocus sequence typing (MLST) <p>Reading:</p> <ul style="list-style-type: none"> ▪ https://pubmlst.org/general.shtml ▪ Multi-locus sequence typing: a tool for global epidemiology Urwin et al, 2003;11:479.

Week 10	<p>Topics: Molecular Epidemiology Tools: cgMLST</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Core genome multilocus sequence typing (cgMLST) <p>Reading:</p> <ul style="list-style-type: none"> ▪ chewBBACA: A complete suite for gene-by-gene schema creation and strain identification Silva et al, Microb Genom, 2018;4:e000166. ▪ https://github.com/B-UMMI/chewBBACA
Week 11	<p>Topics: Molecular Epidemiology Tools: cgSNP</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Core genome multilocus single nucleotide polymorphism <p>Reading:</p> <ul style="list-style-type: none"> ▪ Core genome multilocus sequence typing and single nucleotide polymorphism analysis in the epidemiology of Brucella melitensis infections Janowicz et al, Journal of Clinical Microbiology, 2018;56:e00517. ▪ From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline by Auwera et al, Curr Protoc Bioinformatics, 2013;43:11.10.1. ▪ The In Silico Genotype (ISG): an open-source pipeline to rapidly identify and annotate nucleotide variants for comparative genomics applications by Jason et al, bioRxiv https://doi.org/10.1101/015578, 2015.
Week 12	<p>Topics: Minimum spanning tree with PHYLOViZ</p> <p>Lectures:</p> <ul style="list-style-type: none"> ▪ Minimum spanning tree (MST) based on cgSNP <p>Reading:</p> <ul style="list-style-type: none"> ▪ PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. ▪ http://www.comparingpartitions.info/?link=Tool
Week 13	<p>Topics: Phylogenomic tree construction</p> <ul style="list-style-type: none"> ▪ Phylogenomic tree building methods (Maximum-likelihood, Bayesian, Distance-based) Reading: ▪ BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data by Gascuel O, Mol Biol Evol, 1997;14:685. ▪ http://www.iqtree.org/ ▪ https://www.beast2.org/

Week 14	Topics: Antibiotic Resistance Database Lectures: <ul style="list-style-type: none"> ▪ The Comprehensive Antibiotic Resistance Database (CARD) Reading: <ul style="list-style-type: none"> ▪ https://card.mcmaster.ca/
Week 15	Topics: Antibiotic Resistance Detection Lectures: <ul style="list-style-type: none"> ▪ Resistance Gene (RGI) for prediction of resistome based on homology and SNP models Reading: <ul style="list-style-type: none"> ▪ https://card.mcmaster.ca/
Week 16	Topics: Interactive Tree of Life (iTOL) Lectures: <ul style="list-style-type: none"> ▪ How to visualize and annotate phylogenomic trees with antibiotic resistance profile and epidemiological data from electronic health record to improve our understanding of healthcare-associated infections (HAI) Reading: https://itol.embl.de/
Week 17	Topics: Final Reports, Class Q&A Session