San José State University
Computer Science Department
CS 286: Next Generation Sequencing & Genome Assembly
Fall 2021

Course Information

Course and Contact Information

Instructor:	Leonard Wesley
Office Location:	MH 212
Telephone:	408.924.5287
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	TBD
Class Days/Time:	TBD
Classroom:	TBD
Prerequisites:	Biol 135A or equivalent, CS146 or equivalent, Math 161A or equivalent, and be comfortable running UNIX, Windows, or Mac based application software, or instructor consent.

Course Description

Next generation sequencing (NGS) is a high-throughput method used to determine a portion of the nucleotide sequence of a biological organism's genome. NGS techniques utilize DNA and RNA sequencing technologies that are capable of processing multiple genomic sequences in parallel. This course will provide the student with a thorough understanding of the genomic landscape, description of various sequence generation methodologies and technologies (e.g., Illumina, Ion Torrent, Pac Bio, and Oxford Nanopore). The course will also provide instruction on how to perform basic quality control assessment of next generation sequencing data, and how to use next generation sequencing data to perform *de novo* and comparative assemblies of selected genomes and meta-genomes. Students will become familiar with genome annotation techniques, variant calling, and cloud services for bioinformatic analysis of next generation sequencing data.

Learning Outcomes

Upon successful completion of this course, students will:

1. **SLO-1: Intro & Background:** Be familiar with the NGS technology and genome assembly landscape.
2. **SLO-2: Sequencing Technologies:** Know the theory, methodology, and practice of traditional and next generation sequencing technologies to produce short to long reads of genomic material.
3. **SLO-3: Genome Assembly:** Understand the principles and techniques of de novo and reference genome assembly to build a representation of the sequence or genome from which NGS reads were produced.
4. **SLO-4: Read Mapping (a.k.a Alignment):** Learn how to take sequence data (i.e., reads) and the reference sequence for a species, then compare the reads to the reference in a manner to detect small variations in the sequenced sample, such as SNPs and short InDels.
5. **SLO-5: Gene Annotation & Variant Calling:** Learn the process of identifying the locations of genes and all of the coding regions in a genome and determining the function of those genes. Identify and analyze variants in sequence results.
6. **SLO-6: Staying Current:** Know the relationship between genomics and pharmacogenomics, understand the role of bioinformatics and survey the potential applications of next generation sequencing to modern medicine.

Each SLO above corresponds to a learning module that is described in the course calendar below. That is, there are six (6) learning modules that cover the SLOs described above.

Required Texts

Next-Generation DNA Sequencing Informatics, Second Edition

Edited by Stuart M. Brown, *New York University School of Medicine,* Publisher Cold Spring Harbor Laboratory Press, 2015, ISBN 978-1-621821-23-6

NOTE: The field is advancing so rapidly that the above required textbook will be supplemented with more recent publications as appropriate.

Other Optional Reading Material

A Primer of Genome Science, Greg Gibson, Spencer V. Muse, Publisher Sinauer Associates, 2009, Edition #3, ISBN-10: **0878932364** | ISBN-13: **978-0878932368**

Introduction to Computational Biology: Maps, Sequences and Genomes, Michael S. Waterman, CRC Press. (A statistical oriented view of bioinformatics)

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Andreas D. Baxecvanis and B.F. Francis Ouellette, John Wiley & Sons 2nd Ed. (Includes contributions from several authors providing a wide perspective)

Course Requirements and Assignments

Course Logistics

Students should expect to spend approximately nine (9) hours per week (on average) outside of the classroom preparing for and completing the assigned course work. This includes reading papers, viewing videos as appropriate, completing homework and programming exercises, and so forth. The amount of time that a student actually spends will depend on individual skills and the time allocated to the course. The nine (9) hours per week estimate is based on previous experiences of the instructor and students. So please plan and schedule accordingly.

Some of the class will be taught in "flipped" mode where lectures will be available online, and students will be expected to view these lectures **BEFORE** class, as a regular part of their out-of-class work. Classes will concentrate on answering questions and performing exercises that allow students to practice and use the skills, tools and concepts covered in the lectures. Students should consult the Canvas website at least twice weekly and complete assignments by the specified deadlines.

Previously, students have asked for special exception to policies and procedures for this course. An example includes asking the instructor for extra assignments or work to help improve a grade. Even if such a request is reasonable in the opinion of the instructor, no exception will be given to a student unless it can be made available to the entire class, AND does not constitute significant extra work on the part of students, instructors, graders and so forth. Students should have no concern that other students will receive special exceptions that will not be available to the entire class.

NOTE: University policy F69-24 at http://www.sjsu.edu/senate/docs/F69-24.pdf states that "Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading." However, attendance will be required in order to complete and submit many in-class exercises, quizzes, and exams.

Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus.

Classroom Protocol

Instruction will begin at or within several minutes of the official published start time for the course. Please make sure that cell phones, beepers, and texting devices are turned off during the entire scheduled class time. Excessive audible discussions with fellow students is prohibited so that others are not disturbed. If any subject matter is not understood, please do not hesitate to ask for clarification. If an extended response is necessary to remove doubts, then a request to follow up outside of scheduled classroom instruction time might be made.

Quizzes and Exams

There will be three quizzes, one midterm and a final exam all of which will count toward the final grade as specified in the “Grades” section below. During quizzes and exams, communication with other individuals via any means is strictly prohibited without the express permission of the instructor. Violations will be met with the full impact of SJSU’s academic integrity policy and procedures.

Projects

Several life science or genomics-based projects will be described near the start of the course. Projects will involve applying the skills and knowledge learned in the course to the project. Teams of 2-3 students will be formed to work on a selected project topic. Teams will be required to submit a project proposal before starting on a project, and submit a project report along with working code at the end of the course. Individual student scores on a project will be determined by the content and quality of the contribution of each student toward the project. The score on the course project and project presentation will count toward the final grade (percentage wise) as specified in the “Grades” section below.

Reading, Homework, Programming, In-Class Exercises, Participation

Assignments

Graded reading, homework, programming, and class participation and brief course feedback assignments will be given almost weekly, and will count toward the final gradee. There will be 4 In-class Exercise sessions. These will typically involve forming teams of 2-3 students that work on assigned exercises in the classroom. They provide an opportunity to get started on homework programming assignments that are to be submitted on a designated due date. Participation is mandatory, and scores will count toward final gradee.

Computational Resources

Students are required to make sure that they have access to sufficient UNIX, Windows, or Mac based computational resources (e.g., computers and software) to carryout assignments in the course. An attempt to offer the course in a classroom with sufficient computation resources will be made by the department to support classroom instruction and demonstrations. However, students should be prepared to bring their portable laptops to class.

Week and Class Mtg #	Thur	Tue	Module # & Name	TOPIC	Assignment See Canvas For Module & Weekly Assignment Details and Due Dates
Week 1	8/19	8/24	#1 Intro & Background	**8/19:** - Course Intro, Class background survey/skills assessment - What is NGS & genome assembly, what is its purpose? **8/24:** - Historical DNA Sequencing up to publications of HGS - Basic sequencing workflow. - Tools for analyzing sequences - Historical approaches to base identification & Seq Readout	Module #1 Project Teams Formation By Instructor
Week 2	8/26	8/31	#2 Sequencing Technologies	**8/26:** - NGS and assembly overview - 1st Gen Sequencing o 2D-Fractionation & chromatography-based Seq o Maxam-Gilbert Seq. o Gel-based Sanger Seq **8/31:** - ddNTP-based & extension-based sequencing - 1st generation sequencing	Module #2 <DATE> Last Day To Drop Classes
Week 3	9/2	9/7	#2 Sequencing Technologies	**9/2:** - 2nd generation sequencing o NGS Workflow, Cluster Amplification , Chain Reversable Termination, Ligation Seq, Single molecule Seq **9/7:** - NGS Technologies o Illumina, Roche 454, BioSciences, PacBio, IonTorrent, SOLiD. Oxford Nanoppore.	Module #2 Project Proposals Due TBD <DATE> Last Day To Add Classes

4	9/9	9/14	#2 Sequencing Technologies	**9/9:** - **In-Class Exercise 1** Topics Covered **Week-1 to Week-3:** **9/14:** - Phred Score - FASTA & FASTQ file format - Lander-Waterman Assumption - Challenges with Sequencing	Module #2

5	9/16	9/21	#3 Genome Assembly	**9/16:**
				- **Quiz 1 Covers Topics Week 1 thru Week 4**
				- Current and future applications of NGS
				
				9/21:
				- Sequence assembly overview
				- Seq challenges
				  ○ Repeats
				  ○ Assumptions
					Module #3

6	9/23	9/28	#3 Genome Assembly	**9/23:** - SRA DB - De Bruijn Graphs **9/28:** - De Bruijn graph - Contigs - Scaffolds	Module #3
7	9/30	10/5	#3 Genome Assembly	**9/30:** - **In-Class Exercise 2** Topics Covered Week-4 to Week-6 **10/5:** - Challenges with de novo and De Bruijn graph assembly - Correcting errors in contigs	Module #3
8	10/7	10/12	#3 Genome Assembly	**10/7:** - **Midterm (Full period): Covers Topics from Week 1 thru Week 7** **10/12:** - Quality assessment of NGS reads	Module #3

9	10/14	10/19	#3 Genome Assembly	**10/14:** - Trimming and preparing reads for assembly - Handling single and paired-end reads **10/19:** - De novo genome & metagenome assembly ○ SPAdes or Velvet	Module #3

10	10/21	10/26	#3
			Genome
			Assembly
				10/21:
				- **Quiz 2 Covers Topics Week 7 thru Week 9**
				
				- De novo genome & meta genome assembly continued
				  ○ SPAdes or Velvet
				  ○ Short vs long read assembly
				
				10/26:
				- Quality Assessment of assembled genomes
				- Read Mapping continued
				  ○ SAM, BAM, and CRAM file formats
					Module #3

11	10/28	11/2	#4 Read Mapping	**10/28:** - Read Mapping (cont.) **11/2:** - **In-Class Exercise 3** Topics covered from week 7 to week 10	Module #4
12	11/4	11/9	#4 Read Mapping	**11/4:** - Read Mapping (cont.) **11/9:** - Gene Annotation	Module #4
13	11/11	11/16	#5 Gene Annotation & Variant Calling	**11/11:** - **Quiz 3 Covers Topics Week 9 thru Week 12** - Gene Annotation **11/16:** - **In-Class Exercise 4** (Work on Team Projects, Q&A) Hand out read data sets to assemble. Students start working on assembling the reads.	Module #5

14	11/18	11/23	#5 Gene Annotation & Variant Calling	**11/18:** - Gene Annotation continued **11/23:** - SNP and Variant Calling	Module #5
15	11/25	11/30	#5 Gene Annotation & Variant Calling #6 Staying Current	**11/25** - **Thanksgiving Holiday** **11/30:** - SNP and Variant Calling - Personalized medicine	Module #5 Module #6
16	12/2	12/7	#6 Staying Current	**12/2:** - Personalized medicine **12/7:** - Preparation for Final Project Submission. - Q&A	Module #5
			Final Project Code and Project Report Due To Canvas **TBD** **Final Exam** **TBD**		

SCHEDULE FOOTNOTES:
NONE AS OF Fall 2021

Grades *

WRITTEN HOMEWORK (4 at 10 points each)	40 pts
QUIZZES (3 at 40pts each)	120 pts
MIDTERM	100 pts
IN-CLASS EXERCISES (4 at 50pts each)	200 pts
WEEKLY COURSE FEEDBACK (12 at 5pts each)	60 pts
PROGRAMMING ASSIGNMENTS (2 @ 40pts each)	80 pts
FINAL EXAM & PROJECT REPORT & CODE	400 pts
Total Course Points	= 1,000 pts Total

* The total points for each category might change depending on the number of project teams and assignments. The instructor reserves the right to adjust, with sufficient advanced notice, the above point distribution by ± 5 pts. Such adjustments might be based on the difficulty or simplicity of assignments or quizzes or exams.

Grading Information

Grading Percentage Breakdown

(NOTE: Ranges might change if point totals change)

Grading Percentage Breakdown		
Percent of Total Points	**Points**	**Letter Grade**
96.66%	≥ 967	A plus
93.33%	≥ 933	A
90.00%	≥ 900	A minus
86.66%	≥ 867	B plus
83.33%	≥ 833	B
80.00%	≥ 800	B minus
76.66%	≥ 767	C plus
73.33%	≥ 733	C
70.00%	≥ 700	C minus
66.66%	≥ 667	D plus
63.33%	≥ 633	D
60.00%	≥ 600	D minus
59.99%	< 600	F

HOW TO CALCULATE/ESTIMATE YOUR GRADE
If students would like to calculate their numeric grade percentage, the formula is as follows:

Numeric CS 286 Grade Percentage =

$$\frac{Total\ points\ from\ assignments}{Total\ course\ points}\ x\ 100\%$$

There is no guarantee that grades will be curved. If so, it will be done at the end of the semester. The instructor is already aware that graduate students need to maintain an overall GPA of B or better. Just because a student NEEDS a particular grade doesn't mean that the instructor will automatically GIVE the student that grade. Students must EARN a passing grade based on submitted and evaluated course work.

Extra Credit Options, If Available
There are no extra credit assignments in this course except for completing designated "Advanced" assignments. However, homework assignments and exams might contain extra credit options.

Penalty For Late Or Missed Work
Late assignments will receive a 25% deduction for every 24hr period the submission is late. There will be partial credit for assignments.

Receiving An Incomplete (I) Grade
Receiving a grade of incomplete (I) is not automatic. Students must complete at least 80% of course assignments by the end of the semester to be eligible to receive a grade of incomplete. Students must also provide documentation to support the reason for the request to receive an incomplete grade. The instructor reserves the right to make a final decision regarding giving an incomplete grade. If the instructor agrees to give a student an Incomplete grade, the instructor will enter the remaining work to be completed as part of the PeopleSoft grade submission process.

Grade Change Policy
It is a university policy that course grade changes must be made within one semester from the end of the course. Requests for exceptions to this policy must be accompanied with a documented and compelling reason.

University Policies
General Expectations, Rights and Responsibilities of the Student
As members of the academic community, students accept both the rights and responsibilities incumbent upon all members of the institution. Students are encouraged to familiarize themselves with SJSU's policies and practices pertaining to the procedures to follow if and when questions or concerns about a class arises. See University Policy

S90–5 at http://www.sjsu.edu/senate/docs/S90-5.pdf and University Grading System Policy F18-5 at http://www.sjsu.edu/senate/docs/F18-5.pdf. More detailed information on a variety of related topics is available in the SJSU catalog, at https://catalog.sjsu.edu/ . In general, it is recommended that students begin by seeking clarification or discussing concerns with their instructor. If such conversation is not possible, or if it does not serve to address the issue, it is recommended that the student contact the Department Chair as a next step.

Dropping and Adding
Students are responsible for understanding the policies and procedures about add/drop, grade forgiveness, etc. Refer to the current semester's Catalog Policies section at https://catalog.sjsu.edu/content.php?catoid=2&navoid=98. Add/drop deadlines can be found on the current academic year calendars document on the Academic Calendars webpage at https://www.sjsu.edu/provost/resources/academic-calendars/index.php/ . The Late Drop Policy is available at https://www.sjsu.edu/aars/forms-resources/late-drops.php **.** Students should be aware of the current deadlines and penalties for dropping classes. Information about the latest changes and news is available at the Advising Hub at http://www.sjsu.edu/advising/.

Consent for Recording of Class and Public Sharing of Instructor Material
University Policy S12-7 at http://www.sjsu.edu/senate/docs/S12-7.pdf, requires students to obtain instructor's permission to record the course and the following items to be included in the syllabus:

- "Common courtesy and professional behavior dictate that you notify someone when you are recording him/her. You must obtain the instructor's permission to make audio or video recordings in this class. Such permission allows the recordings to be used for your private, study purposes only. The recordings are the intellectual property of the instructor; you have not been given any rights to reproduce or distribute the material."
    - It is suggested that the Syllabus include the instructor's process for granting permission, whether in writing or orally and whether for the whole semester or on a class by class basis.
    - In classes where active participation of students or guests may be on the recording, permission of those students or guests should be obtained as well.

- "Course material developed by the instructor is the intellectual property of the instructor and cannot be shared publicly without his/her approval. You may not publicly share, or upload instructor generated material for this course such as exam questions, lecture notes, or homework solutions without instructor consent."

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/ and Syllabus Information web page at https://www.sjsu.edu/curriculum/courses/syllabus-info.php. Make sure to review these policies and resources.

Academic Integrity

Your commitment, as a student, to learning is evidenced by your enrollment at San Jose State University. The University Academic Integrity Policy S07-2 at http://www.sjsu.edu/senate/docs/S07-2.pdf requires you to be honest in all your academic course work. Faculty members are required to report all infractions to the office of Student Conduct and Ethical Development. The Student Conduct and Ethical Development website is available at http://www.sjsu.edu/studentconduct/.

Campus Policy in Compliance with the American Disabilities Act

If you need course adaptations or accommodations because of a disability, or if you need to make special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. Presidential Directive 97-03 at https://sjsu.edu/president/docs/PD_1997-03.pdf requires that students with disabilities requesting accommodations must register with the Accessible Education Center (AEC) at http://www.sjsu.edu/aec to establish a record of their disability.