|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |   **Bioinformatics****Winter term, 2014-2015****Instructor:** Christopher Chabot, MSC#64, 535-2864, chrisc@plymouth.edu. **Office/Lab hours:** as arranged. Please feel free to stop by anytime; I have an "open-door" policy.**Meeting times:** MTW – 9-11:30 and as arranged (for work on Projects described below)**Book required:**Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools, 2014, [S. Choudhuri](http://www.amazon.com/Supratim-Choudhuri/e/B00K1I67XA/ref%3Ddp_byline_cont_book_1), Elsevier. $48 (Amazon).**Introduction**We will take a two pronged approach to this course: 1) The heart of this course is in analyzing your own datasets. You can read and talk about bioinformatics, but to really learn it, you have to *do* it. A goal at the outset is to have you, the student, focus on a set of bioinformatics problems that will hopefully form the basis of a thesis chapter or two. Thus, the instructor will meet with each student at the beginning of the course and they will, together, outline a workplan (“project”) that will be expected to be completed during this term. The final **Project** is your chance to use what you have learned along the way and apply it to a topic of your interest. You need to design your project and describe how you expect it to be completed.2) The course will consist of a series of lectures and “workshops”: we will meet officially as a group approximately eight-ten times to cover the “background” underlying principles of the topics below. These topics will be covered by myself and by guest lecturers. In between, you will be expected to work on your projects. B117 will be open to you and will have the software that you need in order to complete your projects/analyses. The end project will be a poster that each student will put together that is comprised primarily of data figures, hopefully suitable for direct inclusion in your thesis. I will ask that you present these results to the class and, perhaps, to your advisor as well.🡪 Chapters 1-4 will primarily be an overview while the rest of the course (and Chapters) will be of a more practical application.To all: **Problems** will occur. Links to specified sites may change. A page may not load properly. Unique problems may surface associated with specific server access. An analysis application may crash or give consistent error messages. Whatever it is, please communicate - both with me and with your fellow classmates. If the problem is associated with a campus computer, or if you think it may be due to the CLC server, or even if you are not sure, try the help line: **5-2929**. **Keeping a log:** It is easy to get caught up in surfing the web, but how do you keep track of all the places you'd like to come back to? How do you minimize the time spent and maximize the information retrieved? How can you be sure that you recorded the URLs [Internet addresses] correctly? How can you speed up the process? The answer to all of the above questions is to make and use a log.A log is an incredibly handy tool and quite easy to set up. Open a page in your favorite word processing program or notepad. [Try using a recent version of Word, because you can automatically turn URLs into active links by using return/enter at the end of the address.] Size the window so you can easily click on it whenever you want, while saving most of your screen for the open web browser window. Stagger the corners, so you can easily toggle between the two windows. Alternatively, you may prefer to work with both windows maximized and toggle between them by using the navigation bar.When you find something you want to save, be it a single address or a whole page of information, you can simply copy/paste between the browser and your log. You can add your own notes and comments as you go, note questions you have, ideas you want to follow in the future, and so on. Be sure to save your log and to back it up!**TENTATIVE SCHEDULE:****Jan. 5 - Chapter 1: Fundamentals of Genes and Genomes**1. Discussion of Chapter 1

DNA – Eukaryotic gene structure (vs (prokaryotic)RNA – coding vs. non codingProtein structure and functionGenome structure – promotors, etc., epigenesis1. Choosing/assigning technique presentations for each student. The below represents an incomplete list. Ideally, each student will present on two of the below techniques.
2. BLASTs -

a) against databases and b) against “local” databases c) initial query blasting (of a published sequence against your database) and d) reciprocal blasting (taking resulting sequence and blasting against public databases) KEY: how to determine if partial sequences in your database are partial due to insufficient data or a problem in search methodology.1. Alignments - Review of alignment tools, which to use and when, is interesting. T-Coffee, may be superior to Clustal Omega, Muscle, etc. (their M-Coffee function compiles these methods with others to give a broader consensus.) [T](http://tcoffee.crg.cat/apps/tcoffee/index.html)-Coffee
2. Alternative splicing - How to ID isoforms vs. gene duplications in database
3. Phylogenetic trees
4. BLAST2GO - JORDAN
5. RNAseq analyses
6. Haplotype analysis
7. Databases and types
8. Plasmid mapping and primer design
9. Annotations
10. SNP analysis???
11. Protein Structure Prediction
12. Metabolism and Networks
13. R stats

 **- Chapter 2: Fundamentals of Molecular Evolution**DarwinismMolecular mutationsRecombinationGene flow, diversityNew gene originsMolecular clock and phylogenetics**Jan. 6 - Chapter 3: Genomic Technologies** Sanger to Next-Gen and Next-next Gen Sequencing **- Chapter 4: A Brief History of Bioinformatics** **Jan. 7 - Chapter 5: Databases & Queries**An introduction to useful sites related to bioinformatics and supporting subjects and instruction on how to go about finding sites of interest by using search engines. Developing efficient search strategies early will help in many ways throughout the semester, and beyond. Databases, including accessing and searching, are introduced next, beginning with literature databases. Molecular sequence database interfaces, such as NCBI, EBI, and NBIF, ENSEMBL are introduced along with using different types of search strategies, which includes locating specific nucleic acid and protein sequences and searching for homologous sequences. Analysis of the quality of search results and how to refine the searches is an important part of this introduction. Other types of databases are explored, including ones related to protein structure, genomics, and metabolic pathways. ***Assignment: Choose a database, explore it, and report back tomorrow on an aspect of the database that you think would be useful to 1) you and 2) another classmate***CLC Biology Workbench, a multipurpose interface with server-based project file storage, is introduced and is expected to be useful for many activities during the semester and beyond. **January 12 - Chapter 6: Genomics and Transcriptomics**Next generation SequencingBLASTing and Alignments, analyzing raw sequence data, including editing and assembling fragments into contigs, aligning and displaying contigs.Also, practice in finding and annotating genes. Further exploration of databases includes becoming familiar with genomic databases associated with many of the genome projects [*Arabidopsis*, *Bacillus subtilis*, *C. elegans*, and human, to name a few] and using applications associated with them. **January 13 - Chapter 7: Genome Sequences (short because few have these data)**Gene prediction and annotation – introns, exons, promotors, etc.Microarrays and the their analysis will be introducedPlasmid mapping and designing primers. **January 14 - Chapter 8: Protein Sequences**Protein Structure PredictionBonds and shape, Hydrophobicity, antigenicity plots**U**se of SwissProt and PDB databases. Motif and profile analysis are extended to examine relationships to structural patterns. 3D structure viewers are used to view protein models and homology modeling is introduced. Secondary and tertiary structure prediction is examined using a variety of tools. **January 20 - Chapter 9: Phylogenetics**We will focus on multiple sequence alignment [MSA] tools, such as ClustalW, and related applications. The effects of selection versus genetic drift is examined using closely related sequences. Sequence profiles to identify and describe motifs and motif databases are introduced. This includes NCBI's COG [Cluster of Orthologous Groups] database, which compares complete genomes and can be used to identify gene families and phylogenetic patterns. Phylogenetic tree building and tree evaluation is introduced and cladistic methods of parsimony, Bayesian, and maximum likelihood are examined. **January 21 - Gene Ontology using BLAST2GO** **Intro to R Stats****January 22: Metabolism & Networks****An** exploration of metabolic databases such as KEGG, an encyclopedia of linked databases of genes, genomes, and metabolic pathways. Data mining methods as applied to biological databases are introduced, including self-organizing maps, decision trees, and neural networks. The brief introduction given here should give an idea of some exciting areas which are expected to grow very rapidly, especially as more fully sequenced genomes become available. Additional Policies:Academic Integrity:[https://www.plymouth.edu/undergraduate/files/2010/11/Academic -Integrity-Policy.pdf](https://www.plymouth.edu/undergraduate/files/2010/11/Academic%20-Integrity-Policy.pdf) Disabilities: Plymouth State University is committed to providing students with documented disabilities equal access to all university programs and facilities. If you think you have a disability requiring accommodations, you should immediately contact the PASS Office in Lamson Library (535-2270) to determine whether you are eligible for such accommodations. Academic accommodations will only be considered for students who have registered with the PASS Office. If you have a Letter of Accommodation for this course from the PASS Office, please provide the instructor with that information privately so that you and the instructor can review those accommodations.• ATTENDANCEAttendance is nearly mandatory and is strongly encouraged in any case. The classes are 2.5 hours during which we will mix both lab and lecture content on a regular basis. Missing one class would be like missing a whole week! There is virtually always a direct correlation between a students' attendance and their grade. Furthermore, we are all depending on your input to the class in order to make it a more interesting class AND class attendance and participation are part of your grade.<https://www.plymouth.edu/undergraduate/files/2010/11/Class-Attendance-Policy.pdf> • CLASS GRADING: PSU Policy: <https://www.plymouth.edu/undergraduate/files/2010/11/Fair-Grading-Policy.pdf> **Grading: grading will be based on** 1. Attendance – 20%
2. Course participation - 30%

-Electronic submission of a log-Oral comments and questions during class-Short presentations of techniques that are instrumental to your final project – these will be assigned/selected the first week of class. Includes database assignment.3) Final project and presentation– 50% **SHORTHAND SYMBOLS FOR AMINO ACIDS**[1-letter symbols are commonly used in sequence data]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **One letter** |  | **Three letter** |  | **Amino Acid** |
|  |  |  |  |  |
| A |  | Ala |  | Alanine |
| R |  | Arg |  | Arginine |
| N |  | Asn |  | Asparagine |
| D |  | Asp |  | Aspartic acid |
| B |  | Asx |  | Asn or Asp |
| C |  | Cys |  | Cysteine |
| Q |  | Gln |  | Glutamine |
| E |  | Glu |  | Glutamic acid |
| Z |  | Glx |  | Gln or Glu |
| G |  | Gly |  | Glycine |
| H |  | His |  | Histidine |
| I |  | Ile |  | Isoleucine |
| L |  | Leu |  | Leucine |
| K |  | Lys |  | Lysine |
| M |  | Met |  | Methionine |
| F |  | Phe |  | Phenylalanine |
| P |  | Pro |  | Proline |
| S |  | Ser |  | Serine |
| T |  | Thr |  | Threonine |
| W |  | Trp |  | Tryptophan |
| Y |  | Tyr |  | Tyrosine |
| V |  | Val |  | Valine |
| **Amino acids with nonpolar [hydrophobic] groups** | **Amino acids with uncharged polar groups [pH = 6-7]** | **Amino acids with acidic groups[pH = 6-7]** | **Amino acids with basic groups[pH = 6-7]** |
|  |  |  |  |
| Ala | Asn | Asp | Arg |
| Ile | Cys | Glu | His |
| Leu | Gly |  | Lys |
| Met | Gln |  |  |
| Phe | Ser |  |  |
| Pro | Thr |  |  |
| Trp |  |  |  |
| Val |  |  |  |

 |
|  |  |