This review sheet is intended to help guide you on both *what* and *how* to study for the upcoming exam.

When/What/How

When is the exam? The exam will be available on Moodle starting at 3pm on Wednesday, April 22nd. It is due by **3pm on Friday, April 24th**. No late submissions will be accepted.

I recommend that you study before you take the exam (like you normally would) and allocate no more than 2 hours to take the exam. You are given a large window to allow for flexibility. You must upload your solutions to Moodle when you are done.

What material can I use on the exam? This exam is open notes, open book, open internet. You may use a calculator. You may NOT discuss the exam with anyone else until after the deadline, even if they tell you that they have already completed the exam.

How can I study for the exam? Here are a few suggestions on how you might most effectively use your study time.

- Go back through the homework problems. If you missed points on something, make sure you understand why.
- Go back through any notes or worksheets from class. Make sure that you understand the inputs/outputs of any algorithms discussed in class/in video lectures and are able to apply the algorithm to small examples.
- Review material from the lectures and re-read any sections of the book that pertain to material you find confusing. The online textbook book does a great job of explaining many concepts. Now that you've tried your hand at some things, the reading may make more sense the second time through.

Material on the Exam

The exam will cover all material covered in the readings, in-class lectures, video lectures and on the homework from the last exam up through clustering. The following table gives a more specific breakdown of both topics you should be familiar with, but also some tasks that you should feel comfortable doing on the exam. Note, there is no guarantee that this list is exhaustive, nor should you expect that each of these will show up on the exam.

Topic	Terms/Topics to know	You should be able to ...
Assembly	Overlap-Layout-Consensus Assembly, de Bruijn graph, de Bruijn Assembly, Hamiltonian path, Eulerian cycles/paths, N50, L50, biological and technical challenges to assembly	(1) Compute N50 or L50 for a given assembly. (2) Explain challenges to assembly and how they may affect produced assemblies. (3) Construct and use overlap or de Bruijn graphs (find Hamiltonian/Eulerian Paths). (4) Compare different assembly approaches
Phylogenetics	phylogenetic tree, distance based and character based phylogeny, additive matrix, additive distance algorithm, neighbor-joining algorithm, perfect phylogeny algorithm, hierarchical clustering, four point condition, Fitch's algorithm, Sankoff's algorithm, phylogenetic distances	(1) Determine if a tree can be constructed from a distance matrix or a character-state matrix. (2) Be able to apply a phylogenetic tree construction algorithm to a small example. (3) Given a new, but similar tree construction algorithm, be able to apply it and answer questions about it. (4) Describe which phylogenetic tree reconstruction algorithms are most appropriate to a given dataset.
Clustering	Hierarchical Clustering, min-link, average-link and centroid approaches	(1) Be able to cluster a given dataset and choose an appropriate way of computing distance between clusters. (2) Use or analyze the output of heirarchical clustering.