Note: There will be 8 assignments throughout the term. Each assignment is worth 8% of your final grade. Your lowest assignment grade will be dropped providing a total for the assignment component of your final grade as 56%.

Note: Please adhere to the stated expectations: "I expect that you'll interact with your colleagues throughout the course, in discussing any readings, during class, or in considering any assignments. However, your submitted assignments should reflect and be of your own individual work."

Note: When searching online you shouldn't need to request custom pricing or quote information. That is, for questions requiring online work, only use information that is readily available online.

*Note: Unless otherwise noted each assignment is due at the *beginning* of the class following its assignment. Please submit your assignment either as hard copy or by email attachment to endy@stanford.edu.*

Note: Please send any questions about this assignment by email to endy@stanford.edu. I'll try to respond in real time.

Part 1. The first part of the assignment is designed to give you at least some familiarity with online databases containing (mostly natural) DNA sequences. Start by finding the DNA sequence associated with NCBI Accession Number NC_002058.

a. What organism is this DNA sequence taken from? (hint: try "Googling" the accession number). ***Poliovirus***
b. How many base pairs comprise this specific DNA sequence? ***7440***
c. Is the DNA sequence part of the organism's genome or the entire genome? ***Entire genome***

Part 2. Next, work through what it would take to construct NC_002058.

a. How many oligonucleotides would you need to purchase in order to assemble NC_002058? (Hint: you will need to define your oligos' average length and oligo-oligo overlap) ***If using oligonucleotides of length 60, with a 20 bp oligo-oligo overlap, then 186 oligonucleotides total, rounding up. You can compute the number of oligos needed by laying out the assembly of oligos across the entire genome (or gene). Mathematically, the resulting formula is: L = length_oligo + (N_oligos - 1)(length_oligo - overlap_oligos). Substituting specific values into this equation for the example here gives: 7440 = 60 + (N-1)(60-20). Solving for N gives 185.5. You can round up to 186, which would result in slightly less overlap for a few of the oligo-oligo pairings.***
b. What would it cost you to order all of these oligonucleotides? (Hint: Integrated DNA Technologies, Inc. is one good custom oligo supplier, but there are many others online). Document your pricing by including a URL to pricing with your assignment. ***Integrated DNA technologies provides good quality oligos. Their website (http://www.idtdna.com/catalog/CustomSynthesisandPurification/Page1.aspx) lists oligo costs at US $0.30 per base pair. So, if you needed to order 186 60bp oligos, that would cost you $18 per oligo, or $3348 total. You could likely get a 50% discount but ordering the oligos in a 96-well plate format (i.e., bulk order).***
c. How much would it cost you to order NC_002058 as a single "gene-length" construct? (Hint: search "gene synthesis"). Be careful to confirm that the custom "gene synthesis" service you choose can deliver a sequence up to the full length of NC_002058. Again,

document your pricing by including a URL. ***Gene synthesis looks to be about US $0.50 per bp today. See http://www.google.com/search?q=gene+synthesis for example. Thus, 74400*0.50 would be $3720.***

d. Based on information available online, how much time do you think that it would take to assemble NC_002058 from oligos? How long do you estimate it would take for a full-length "gene" synthesis order to be delivered? ***You could likely receive the oligos in a few days, perhaps even overnight. Then you would need to implement some assembly method (e.g., PCR assembly) in your own lab. This could take some time. Depending on how good you were at PCR assembly, perhaps you could assemble the full genome in a week, or it might take you a year (or more). Commercial gene synthesis of ~7kb would likely take a few months.***

Part 3. Returning to the NCBI's "Nucleotide" and "Genome" databases:

a. What is the most interesting or important DNA sequence that you can find? Provide the sequence's NCBI Accession Number and sequence length. What would it cost you to construct this sequence? ***Lots of good choices. I like NCBI GenBank Accession Number DQ257435, which leads to the mysterious sounding "Uncultured organism HF10_3D09, complete sequence" file. If you search through this file you can find a gene encoding "proteorhodopsin," which is 836 bp long. Proteorhodopsin is a "simple/primitive" photosynthetic protein in ocean bacteria. It takes sunlight and makes a proton gradient across the cell membrane. You can put the protein in other bacteria (e.g., E. coli) and it does the same thing (i.e., photosynthetic E. coli). What's really amazing is that this protein was only first discovered in 2000, even though it is responsible for much of the photosynthesis on our planet. It would cost $400-500 dollars to have this gene synthesized commercially.***

b. What is the most frightening or dangerous DNA sequence that you can find? As before, provide the sequence's NCBI Accession Number and sequence length. What would it cost you to construct this sequence? ***NC_004161 is the complete genome sequence for a strain of Ebola, a hemorrhagic fever that is often fatal. The genome is 18,891 bp long, which means that it would cost roughly $9500 to have synthesized commercially. Practically, you would likely attract a lot of attention if you tried to order this piece of genetic material.***

Part 4. Read the following pair of one-page editorials.

"Recipe for Destruction" by RAY KURZWEIL and BILL JOY
http://www.nytimes.com/2005/10/17/opinion/17kurzweiljoy.html

"1918 Flu and Responsible Science" by PHILLIP A. SHARP
http://www.scienceonline.org/cgi/content/short/310/5745/17

a. Do you believe that the genome sequences of pathogens should be freely available online? (yes or no) ***Yes (but your answer may vary!)***

b. Why? (please limit your reasoned answer to 250 words or less) ***No disease has been cured in secret. Nature is already producing new emerging infectious diseases. We have to work together to figure out treatments. People will continue to cause harm using biotechnology (e.g., 2001 anthrax attacks). We have to prepare for this, by both minimizing the number of people who would misapply biotech, but also by having effective and fast responses, and a strong public health system.***