Nancy Moran: ECOL 453/553 (Fall 2009)

Oct 27: **Sequencing genomes. Methodological considerations.**

Topics Covered:	Genome sequencing strategies The first sequenced genome: *Haemophilus influenzae* The Human Genome Project Assembly strategies (Hierarchical vs. Shotgun) New sequencing technologies
Readings:	Gibson & Muse, pp. 13-21, 63-69, 79-89, 92-94 Venter et al. (2001) Science 291: only pages 1304-1316 IHGSC (2001) Nature 409: only pages 860-874, 896-901
Assignment:	For Oct 29 -- be ready to discuss Green (2002). For Nov 3: Read your selected paper describing a complete bacterial genome sequence and answer the questions in the assignment below.

Homework Assignment: Due before Class on Nov. 3 2008

BACTERIAL GENOME SEQUENCES

This assignment must be done on your own (*i.e.*, without conferring with others in the class or elsewhere). **Web resources and primary literature can be used but not copied.**

Please download and read your assigned bacterial genome sequence paper, then answer the following questions. [Note that, depending on the particular paper that you read, some of these questions will be trivially easy (i.e., the answer will be mentioned in the Abstract, presented in a Figure or Table, or the subject of an entire section), whereas other questions might require some additional background reading or web-browsing.]

In the following questions, I use the term "your microbe" to refer to the particular bacterial strain or species whose complete genome sequence is the subject of the paper that you read.

1. For your microbe, list the following: (2 points each; **10** points total)
   a. Latin binomial (including strain designation)
   b. The size of its genome (in nucleotides)
   c. Its overall base composition (percentage of guanine plus cytosine residues)
   d. The predicted number of genes in its genome
   e. Does this predicted number of genes include structural RNAs?

2. For your microbe
   a. What is the number of ribosomal RNA operons in its genome? (**2** points)
   b. What is the number of transfer RNA (tRNA) genes in its genome? (**2** points)
   c. In all likelihood, the total number of tRNAs is different from the total number of codons needed to translate a protein. How can you account for the differences in the number of tRNA and the number of codons? (**4** points)

3. What is the stated rationale for sequencing this microbe? What do the authors hope to find out by obtaining its complete genome sequence? (**5** points)

4. Does the genome of your microbe contain any replicating units besides a single circular chromosome? If so, what are these and their sizes, and what do they encode? (**2** points)

5. What are the three closest fully sequenced relatives of your microbe that are currently available in the databases? (The are many ways to get the answer to this question. Please explain how you derived your answer to this question.) (**10** points)

6. Explain in detail, the methodology used to obtain the sequence of this genome. (Do not explain the assembly or annotation process, just the manner in which the DNA was isolated, the library was constructed, the sequenced was done, and the genome sequences. including numbers of reads, clone size ranges, coverage, *etc*. You may need to go back through several years of references to find the actual procedure cited and used). (**15** points)

7. Figure 1 (or some cases, figure 2) in the manuscript is a circular representation of the genome showing its features in a series of concentric circles. For each of these circles, please list the feature that is being mapped and explain *why* a circle is devoted to this particular feature (*i.e*., why it is necessary and useful to have information pertaining each particular property. Please do not simply re-state what the circle represents). (**25** points)

8. What are two of the most interesting findings about your microbe with respect to the *functions* of proteins that are (or those that are not) encoded by its genome? (**10** points)

9. Describe one interesting feature about the *structure* of this genome (as opposed to the functions of proteins encoded by this genome)? Such structural properties might relate to its sequence features, repetitive elements, rearrangements, gene locations and organization, or findings from comparative analyses. (**5** points)

10. How many times has this paper been cited? (To do this, go to the ISI Web of Science on the UA library site). (**5** points)

11. About what do you estimate as the cost (in dollars) to complete the genome sequence of your microbe (including assembly and annotation)? Based on your estimate, do you think that this genome was a "good value"? (Please justify your answer.) (**5** points)

Bacillus cereus	Ivanova et al, 2003, Nature 423:87-91
Bartonella henselae	Alsmark et al, 2004, PNAS 101(26):9716-9721
Burkholderia pseudomallei	Holden et al, 2004, PNAS 101(39):14240-14245
Chlorobium tepidum TLS	Eisen et al, 2002, PNAS 99(14):9509-9514
Clostridium tetani	Bruggemann et al, 2003, PNAS 100(3):1316-1321
Corynebacterium diphtheriae	Cerneno-Tarraga et al, 2003, NAR 31(22):6516-6523
Escherichia coli CFT073	Welch et al, 2002, PNAS 99(26):17020-17024
Ehrlichia ruminantium	Collins et al, 2005, PNAS 102(3):838-843
Francisella tularensis	Larsson et al, 2005, Nature Genetics 37(2):153-159
Lactobacillus acidophilus NCFM	Altermann et al, 2005, PMAS 102(11):3906-3912
Mycobacterium bovis	Garnier et al, 2003, PNAS 100(13):7877-7882
Nocardia farcinica IFM 10152	Ishikawa et al, 2004, PNAS 101(41):14925-14930
Photorhabdus luminescens	Duchaud et al, 2003, Nature Biotechnology 21(11):1307-1313
Salmonella enterica	Chiu et al, 2005, NAR 33(5):1690-1698
Shewanella oneidensis	Heidelberg et al, 2002, Nature Biotechnology 20(11):1118-1123
Shigella flexneri 2a	Jin et al, 2002, NAR 30(20):4432-4441
Streptomyces coelicolor A3(2)	Bentley et al, 2002, Nature 417:141-147
Symbiobacterium thermophilum	Ueda et al, 2004, NAR 32(16):4937-4944
Synechococcus sp. WH8102	Palenik et al, 2003, Nature 424:1037-1042
Staphylococcus saprophyticus	Kuroda et al, 2005, PNAS 102(37):13272-13277
Treponema denticola	Seshadri et al, 2004, PNAS 101(15):5646-5651
Tropheryma whipplei	Bentley et al, 2003, Lancet 361:637-644
Vibrio parahaemolyticus	Makino et al, 2003, Lancet 351:743-749
Wolinella cuccinogenes	Baar et al, 2003, PNAS 100(20):11690-11695
Zymomonas mobilis ZM4	Seo et al, 2005, Nature Biotechnology 23(1):63-68