

## Class 12: Sequence Assembly (1)

**Held:** Tuesday, 6 October 2009

**Summary:** We consider the various techniques that are used to build the large DNA databases that we rely upon.

### Related Pages:

- EBoard.
- Reading: Chapter 5; Markham et al. al. 1998..
- Due: Reading Response: Markham et al..

### Notes:

- We will not be doing a programming project for this chapter. We will, however, discuss the algorithms used for shotgun alignment.
- We're still debating what to do about the protein alignment project. We may return to it on Thursday.
- Any upcoming events?

### Overview:

- Discussion of Markham et al.
- Sequencing DNA

## Discussion of Markham et al.

### Sequencing DNA

- Okay, how do we get all the wonderful sequence data that we've been using, at least for DNA sequences?
- The Sanger method is used for short DNA sequences.
  - The book tells us it's really the only strategy used.
  - We'll look at how it works and some of the data it produces
  - Unlike the book, we'll use Ridom TraceEdit to explore data (TraceEdit is available for all three platforms.)
- Note: You may have wondered why the official FASTA format permits more than A, C, G, T, U, and N for DNA/RNA sequences. Hopefully, the sequencing data we looked at gives you a sense as to why.

Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.