Shotgun sequencing

Sanger DNA sequencing, as you may guess, only works for a certain distance beyond the sequencing primer (best from about 30 nt to 350 nt; the "read length"). Beyond that, very few products are produced because chain termination has already occurred. Therefore, to sequence a longer DNA, special methods are required. An obvious solution is to sequence as far as you can from your primer; then synthesize a new primer near the end of the known sequence; and repeat. Works, but at best you'd be able to sequence maybe 500 bases a day—making it impossible to sequence something like the human genome, with its *billions* of bases.

Another approach, used to sequence very large amounts of DNA (such as an entire genome), is **shotgun sequencing.** In this strategy, the DNA is first shredded into **smaller fragments** which can be **sequenced** individually. The sequences of these fragments are then **reassembled** into their original order, based on **overlaps**, ultimately yielding the complete sequence.

"Shredding" of the DNA can be done using restriction enzymes, or mechanically by shearing. Alignment of the sequences of overlapping pieces is done by computer. In the case of the human genome project, a massive amount of data is involved, requiring supercomputer technology.

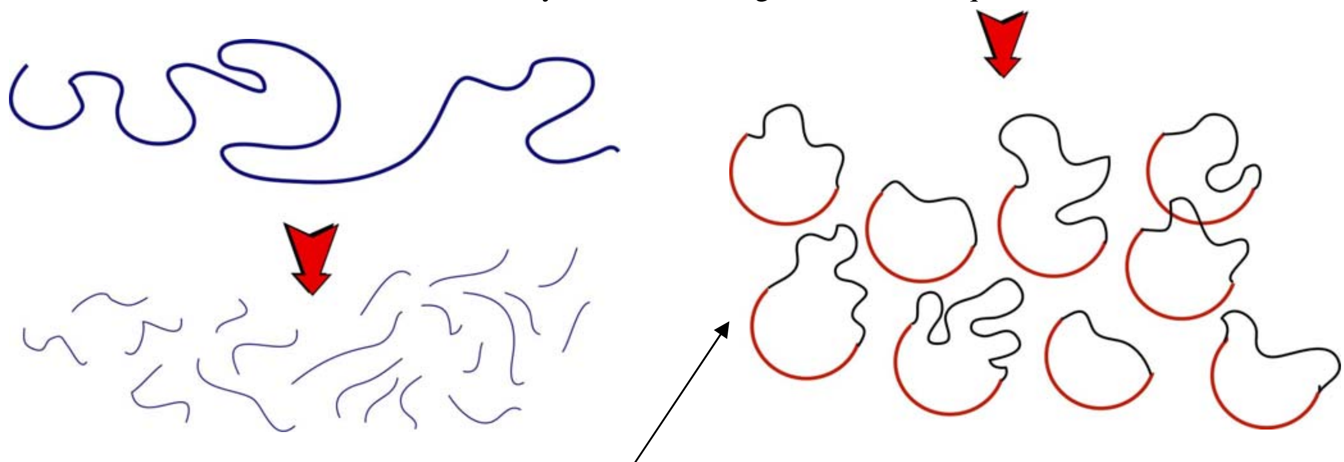
Shotgun Sequencing

To sequence a clone longer than the average read length, it is possible to use a shotgun approach. The idea is to pepper the DNA with sequence reads such that they overlap, and when assembled, yield the complete sequence of the clone.

The shotgun part comes from the way the clone is prepared for sequencing: it is randomly sheared into small pieces (usually about 1 kb) and subcloned into a "universal" cloning vector. The library of subfragments is sampled at random, and a number of sequence reads generated (using a universal primer directing sequencing from within the cloning vector). These sequence reads are then assembled into **contigs**, and the complete sequence of the clone generated.

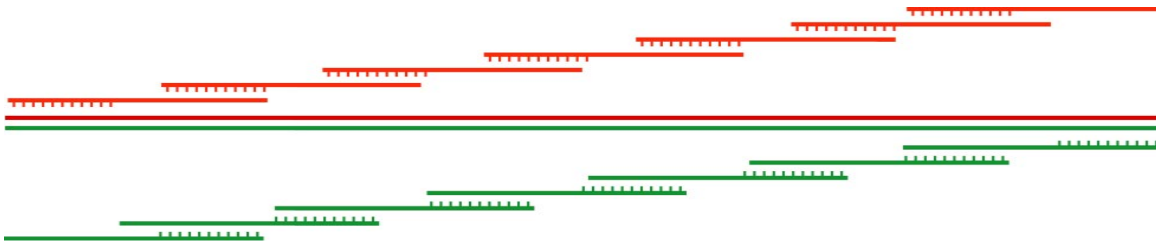
Making a shotgun library

Genomic DNA is sheared or restricted to yield random fragments of the required size.



The fragments are cloned into a universal vector.

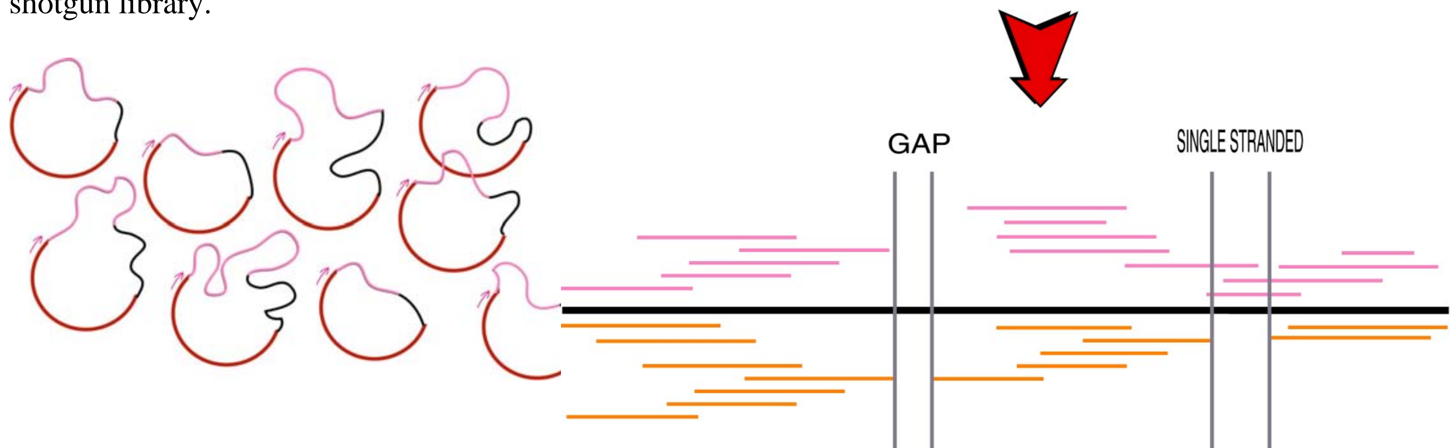
Below is a **contig** (for "contiguous sequence"). The red-green hybrid in the center is the original dsDNA to be sequenced. It was broken up, and smaller pieces cloned into plasmids. The inserts in various randomly chosen plasmids were then sequenced to give the smaller fragments shown. Note that it is important to sequence ***both strands***. While this may seem a waste of effort given the rules of Watson-Crick base pairing, the fact is that certain areas on one strand may be difficult to sequence accurately (for example, because of local secondary structure formation). The complementary strand, however, may sequence well. Using primers from opposite ends will give you sequence for both strands.



Once you have sequenced a bunch of small fragments, a computer can find regions of overlap (shown as hatch marks above) and properly align them into the complete original sequence.

Shotgun Sequencing

Sequencing reactions are performed with a universal primer on a random selection of the clones in the shotgun library.

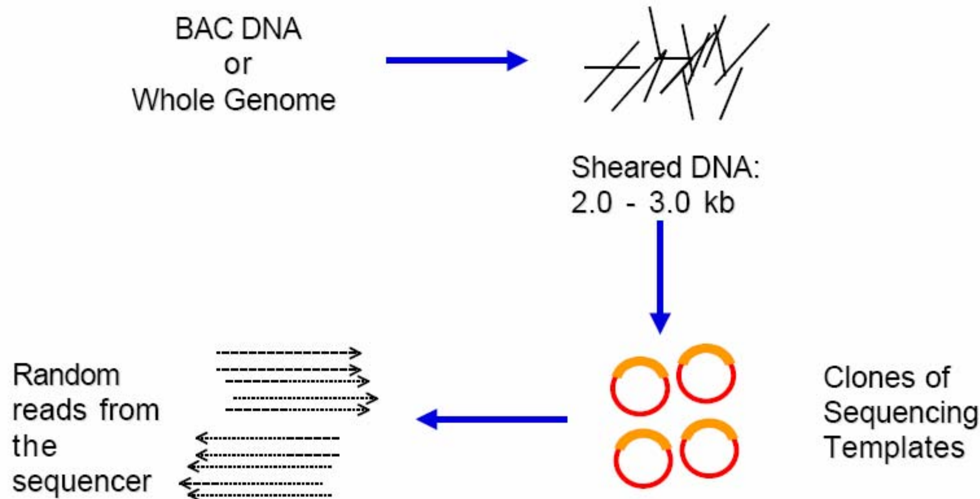


These sequencing reads are assembled in to contigs, identifying gaps (where there is no sequence available) and single-stranded regions (where there is sequence for only one strand).

The gaps and single-stranded regions are then targeted for additional sequencing to produce the full sequenced molecule.

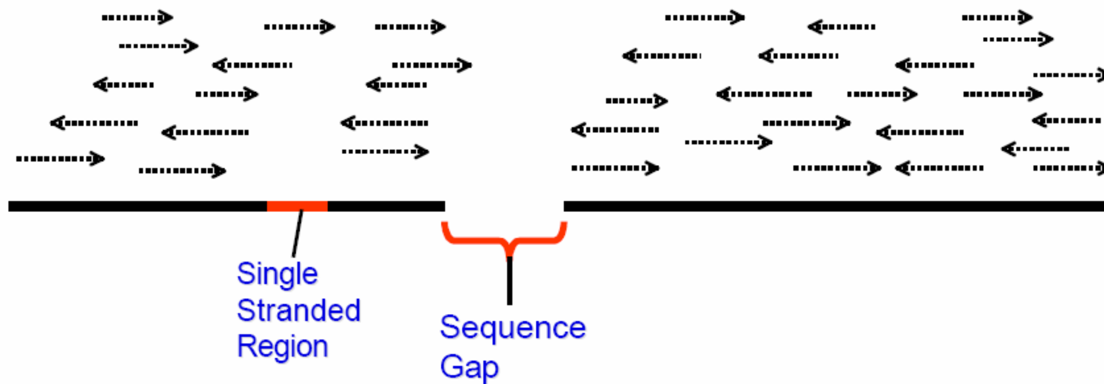
SOURCE: http://www.nematodes.org/teaching/genomics/Genomics3.html

The example of Shotgun Sequencing
I :RANDOM PHASE

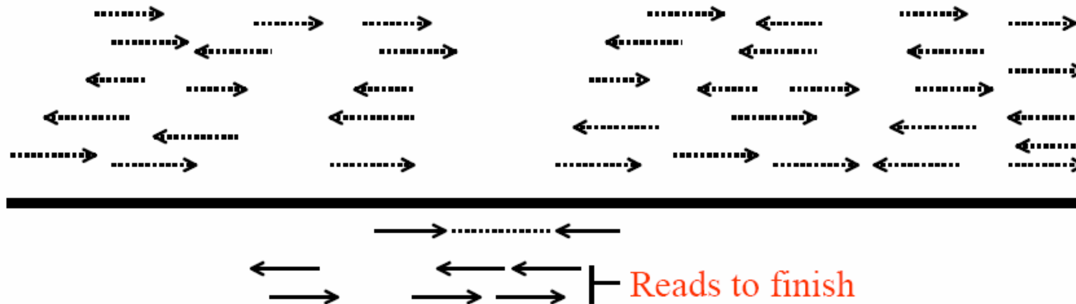


Shotgun Sequencing II:ASSEMBLY

Sequence overlap between individual reads is used to assemble a contiguous set of reads "the contig".



Shotgun sequencing III: FINISHING



- PCR or other methods are used to isolate and sequence the “unfinished regions”.