

Lecture 19

Walking in the genome

Ordered libraries Why walk on the genome? We've talked about methods of handling libraries in which the clones are not organized, or isolated from each other - for example, suppose you have a genomic library of an organism you are studying...

...you might have a tube containing 10^{11} bacteriophage, representing 2×10^5 independent clones, each of which carries an average of 20 kbp inserted. Suppose also that you have a cDNA probe for the coding sequence of "Y", and want to isolate the gene for "Y". Why would we care to have the entire gene? Well, the cDNA might tell us the coding sequence of the gene, but by itself it tells us nothing about the promoter for the gene, the intron/exon structure, or the structure of the chromosome.

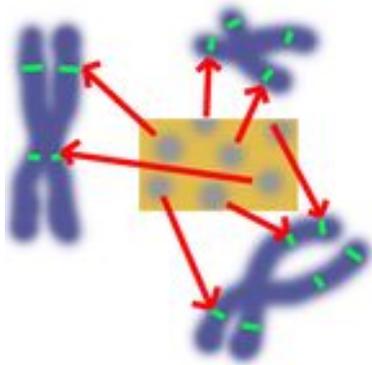
You might take 5,000 plaque forming units (pfu) from the library, mix them with bacteria and plate them on a standard bacteriological plate. At that level, the plaques would be covering the plate, and there would be very few uninfected bacteria. The surface of the plate might look like this, up close, where the yellow color is uninfected bacteria and the clear gray circles are plaques:



You can lift nylon or nitrocellulose filters from the plate to collect an imprint, as we've discussed:

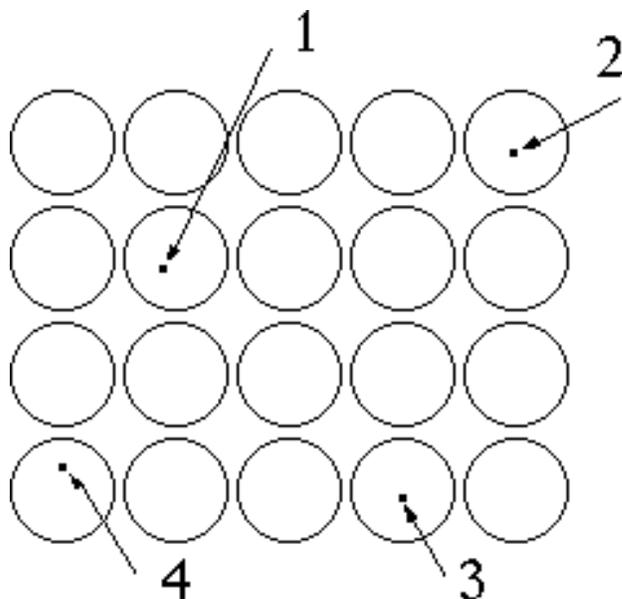


But there the positions of the plaques on the plate are entirely random - it is just a matter of chance where the phage particles end up on the plate and establish a plaque. There is no relationship between where a plaque appears, and what part of the genome it represents. The position of the plaque on a plate and the part of the genome it covers might be represented like this, though of course there is no way to predict the outcome:



That's OK - you can live with that. Here's what you do...

Let us say that you prepare 20 bacteriological plates, each with 5,000 plaques on it, and lift nylon membranes from the 20 plates. You use the cDNA for gene "Y" as a labeled probe, and detect 4 spots among the 20 filters:

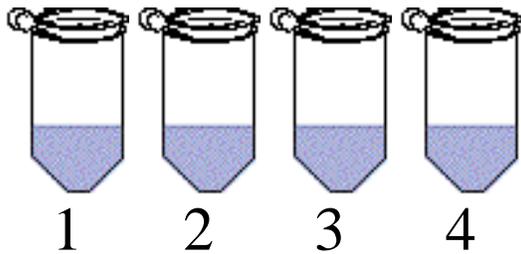


20 Nylon filters

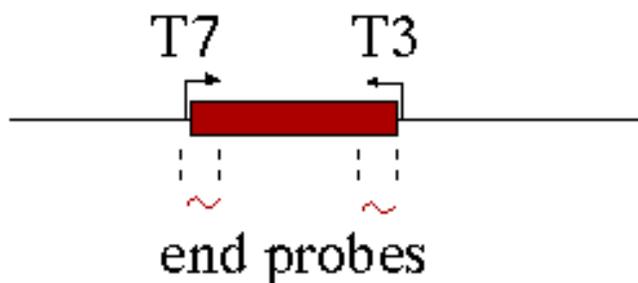
...probed with cDNA for gene "Y"

You go back to the plates from which these filters were lifted, and align the filters so that you can isolate the phage that were

responsible for the hybridization (this takes several more plating experiments to obtain a plaque-pure bacteriophage clone). Now you have four tubes, each representing a different clone:



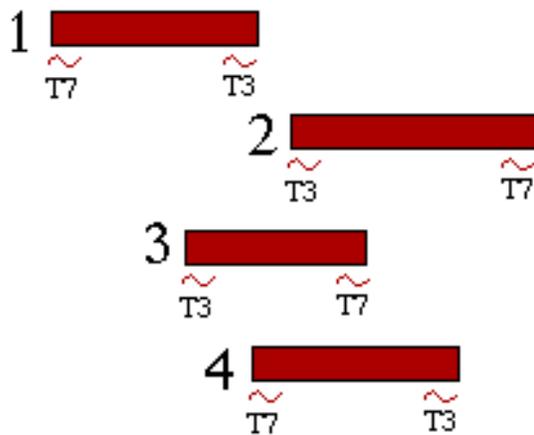
How do you figure out how these four are related to each other? Are they overlapping in sequence? Are they covering a similar sequence but from different loci? Here's a way you can start to organize the collection. The bacteriophage vector has sequences flanking the site of insertion that permit specific bacteriophage RNA polymerases to bind. For example, on one side might be a phage T7 RNA polymerase promoter and on the other might be a phage T3 RNA polymerase promoter. These are different recognition sequences and the bacteriophage enzymes only recognize their own promoter sequences. You can synthesize a bit of RNA at each end of the cloned sequence, using the specific RNA polymerases in vitro and ribonucleotide substrates, and you can radiolabel or chemically label the product. These labeled nucleic acids are called **end probes** because they are at the ends of the clone.



Suppose that you prepare T7 and T3 end probes for each of the four phage (that's eight probes in all) and use these probes in hybridization tests. Maybe you'll get a table that looks like this, where the "+" indicates hybridization between the end probe and target:

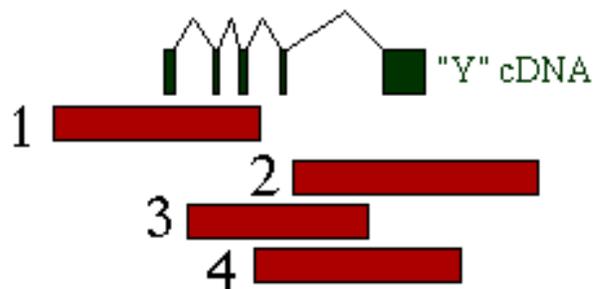
	TARGET PHAGE			
	phage 1	phage 2	phage 3	phage 4
phage 1 T7 probe	+			
phage 1 T3 probe	+		+	
phage 2 T7 probe		+		
phage 2 T3 probe		+	+	+
phage 3 T7 probe		+	+	+
phage 3 T3 probe	+		+	
phage 4 T7 probe			+	+
phage 4 T3 probe		+		+

Such a table would be consistent with the following overlapping map of the insertions in each phage:

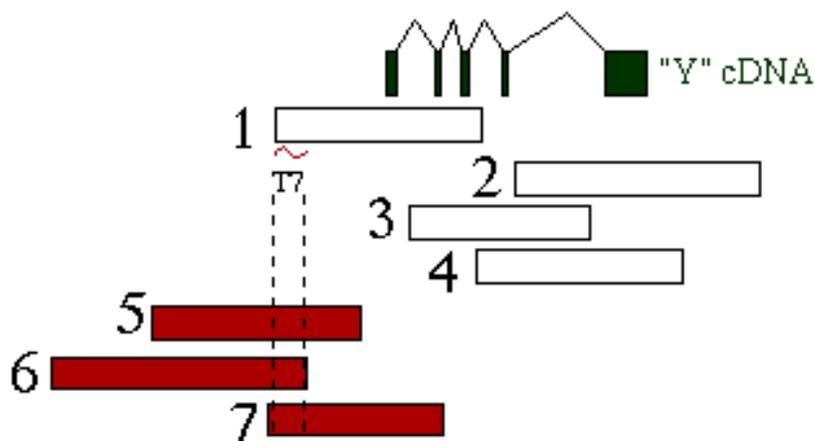


Note that the orientation of the phage insert (i.e. which end is T3 and which is T7) is random.

With this set of four bacteriophage, we may have spanned about 40 kbp. Do we know where our gene "Y" may lie? Well, we know that the cDNA for gene "Y" was able to hybridize to all four phage clones, so it must span clones #1 and #4. Maybe the positions of the exons, with respect to the four clones, would look something like this:

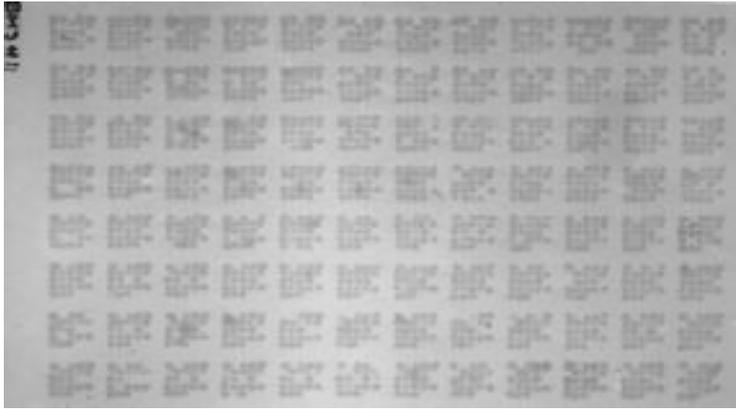


None of the four clones contains the entire gene, but between them the entire gene is represented. If we wanted to obtain additional clones at the 5' end, we could use the T7 end probe from clone #1 to rescreen the library. That way, we can take a step to the left...



Now that we have taken a step, what's to stop us from creating an ordered list of clones? In this example, the order of the left ends would be 6-5-7-1-3-4-2, and they might represent about 60 kbp of a genome. If we were using BAC or PAC libraries, which have larger capacity, the ordered group of clones might represent 600 kbp of the genome.

If you have an ordered library, you have sketched out the structure of the genome, and in many sequencing projects this is the first step in organizing the DNA so that it can be efficiently sequenced. The library is maintained in microtiter dish wells, and may be gridded by a robot for hybridization test. Maybe it would look like this:



The point is that each spot on the filter corresponds to a microtiter well, and each well corresponds to a cloned DNA, and the relative orders of the clones may be largely known.

Problems staying on the track There is a problem with walking on the genome, and that is knowing where you are - that you have not taken a step in the wrong direction.

Suppose for example, that while out for a stroll on the genome you accidentally make an end probe that is in a non-unique sequence. Maybe you've wandered into a gene family and there are many more sequences on other chromosomes that resemble the one you've got. You could take a step to the wrong part of the genome, simply because the probe matched a clone on a different chromosome:



If the repeated DNA were "highly repetitive", you would probably notice the problem before you got too far, because an unexpectedly large number of clones would be positive using the end probe. You need to maintain vigilance, and there are a few methods that can

help.

You can check your position by use of mapping data, particularly against "radiation hybrid maps"

Here are a couple of reading resources:

[Radiation hybrid mapping](#) - Melcher

[Physical mapping of genes in somatic cell radiation hybrids](#) - Lin et al.

You can also establish your position within about 1 to 3 Mb, based on FISH mapping.

[FISH & CHIPS](#) - Max Planck Inst.

Sequence tagged sites can also direct the isolation of overlapping clones:

[A database of mapped human BAC clones](#) - Morley et al.

Stan Metzenberg
Department of Biology
California State University Northridge
Northridge CA 91330-8303
stan.metzenberg@csun.edu

© 1996, 1997, 1998, 1999, 2000, 2001, 2002