Lecture 16

Libraries

A big collection of clones.

What is a library? We know what a "library" is in real life - a room or building filled with a collection of books. In molecular biology we use the term in a slightly different way, to mean a collection of recombinant vectors containing different inserted sequences. You might, for example, clone random genomic DNA or cDNA fragments into a vector, with the hope of identifying a gene you are trying to clone through screening of the library. When random fragments are cloned, the library is sometimes called a "shotgun" approach, because one doesn't really need to aim too carefully to hit the target with a pellet. When a library is designed to contain many possible versions of a sequence, often by synthesis using degenerate oligonucleotides, it is often called a "combinatorial" library.

First let us discuss the vectors that can be used - for example lambda Before we get too involved in discussing the DNA insertions in a library, let us discuss the vectors that hold the foreign DNA. You could think of this as being the "bindings" and "covers" for the books in a real library - they carry the information by holding together the pages.

Almost any recombinant vector can be used to make a library, but some are easier to work with than others. Phage lambda derivatives are easy to use for genomic and cDNA libraries, because the plaques are a stable reference stock for your clones. When you lift a membrane from the collection of plaques, you always leave some plaques behind that can be picked out and propagated.


We have discussed this already for lambda gt11 - the idea that you can grow a "lawn" of bacteria with plaques in it, induce expression of the fusion protein with IPTG, and collect an imprint of the plaques on a circle of nitrocellulose. You could detect your specific phage using an antibody probe (if an expression system) or a DNA probe (by hybridization). For gt11 we'll assume that we've got an antibody probe.

Here's how it works. First, you mix the E. coli host and the lambda gt11 library in a tube, wait about 30 minutes for the phage to adsorb to the bacteria, then mix in some molten agar-containing medium (cooled to 45 C), and "plate" the mixture onto a bacterial plate that already has a layer of solidified agar-containing medium. This first layer of agar may contain X-gal and IPTG (although it is usually the case that the IPTG is soaked into the nitrocellulose filter - we'll forget about that so it is easier to show you what happens)
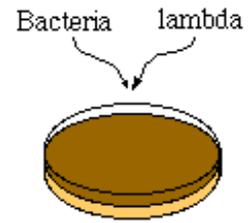
After overnight growth the plate might look like this when you hold it up to the light. Plaques are visible, and some of the plaques are blue. As you recall, the blue plaques represent bacteriophage that do not have insertions, so the lacZ gene is uninterrupted. The colorless plaques have insertions into the lacZ gene. If this is a cDNA library, then we expect that each colorless plaque might represent a different inserted cDNA.
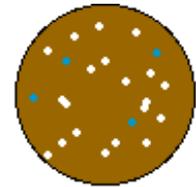
If we apply a nitrocellulose membrane to this plate, allowing it to soak up the proteins that have been expressed and released from the dying cells, then probe the nitrocellulose with an antibody and detect the antibody binding with a secondary antibody and enzyme-linked detection system, we might see results as shown on the right. Note that it is important to align the nitrocellulose membrane with the plate so you know which end is up!
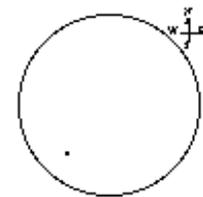
If you have marked the nitrocellulose so that you know its orientation, you can go back to the plate and cut out (punch out) the plaque that generated the positive signal by antibody binding. There will still be some phage in the plaque that can be regrown, although some of the phage are stuck to the nitrocellulose (and lost) and others will have diffused out of the plaque.
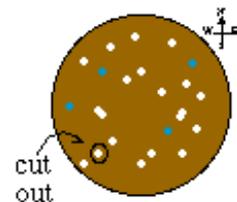
Now you have a phage stock that is partially purified. Since the phage diffuse in the plate, you can only expect that your specific phage will be enriched in the population that you cut out. You need to repeat the plating experiment and detection system several times to enrich for the phage that is the right one. Your first attempt at replating might go something like this:

Note that a fraction of the plaques on the plate are positive with the antibody, but most are still other things. You then need to pick an isolated "positive" plaque from this second generation plate and try again with a third (and probably fourth) generation plate to increase the level of purity.
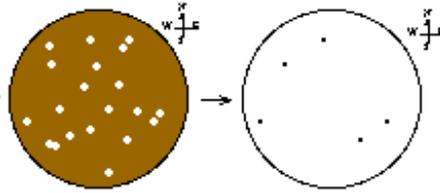
This is what you eventually hope to see. In this case, every plaque in the sample is detected as a "positive" with the antibody probe. That means that the phage you have picked is now "plaque purified", and all of the phage in the stock are of one kind. It is now a clone.
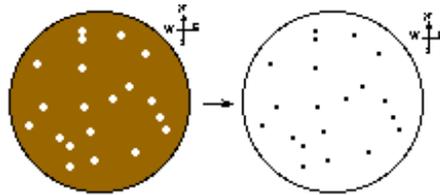
A lambda clone can be grown in large quantities and stored as a purified bacteriophage stock, at a concentration of 10^9 to 10^{11} phage per ml. You may add chloroform to the stock to preserve it. You may also preserve the lambda clone as purified lambda DNA, but to infect cells you would need to "package" the DNA into capsids again.

What next?

You now have a gt11 clone that expresses a fusion protein that reacts or cross-reacts with your antibody probe. It could be a good result, but then again it might not be the protein you are really looking for. You will need to sequence the inserted DNA to see what you've got, and it is often the case that you want to move the inserted DNA to a plasmid vector.

If the EcoRI sites are still intact in the gt11 vector, you could isolate the insertion by digesting the purified lambda DNA with EcoRI and isolating the foreign DNA insert on a gel. You could also use the polymerase chain reaction to isolate the insertion, but this might be difficult if the insertion is large (remember that gt11 can clone up to 9 kbp).

lambda ZAP Lambda gt11 has been around for quite a while now, and there are more elaborate versions called lambda ZAP vectors:

Lambda ZAP II vector

http://www.stratagene.com/figures/Lambda_ZAPII.gif
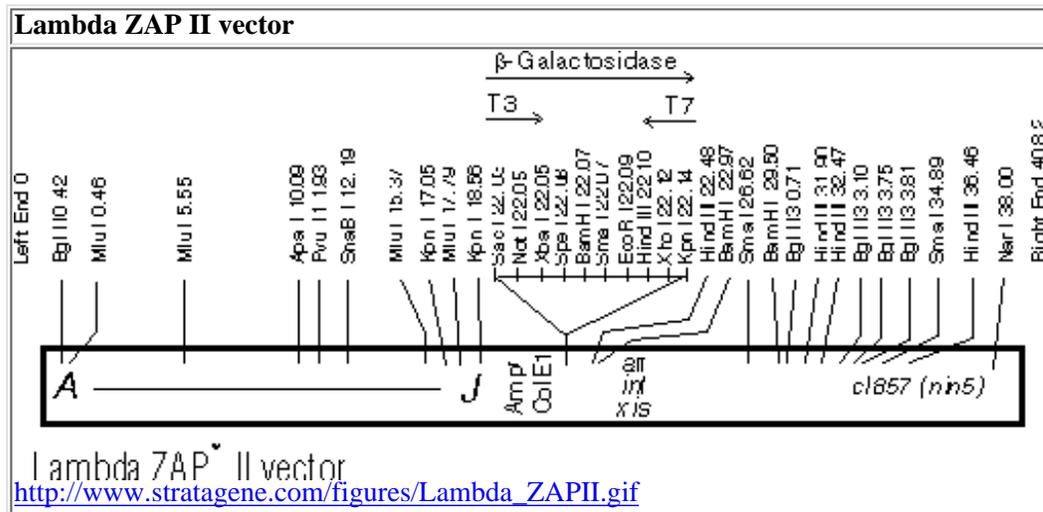
Note that a polylinker has now been inserted into the lacZ gene - you now have a choice of 10 cloning enzymes! There are also T7 and T3 RNA polymerase promoters surrounding the polylinker, an ampicillin resistance gene, a ColE1 origin, an f1 phage origin of DNA replication and terminator of replication (not shown in map).

You can "subclone" the middle of the phage automatically!

From what we've discussed so far, you might consider using enzymes that cut the polylinker to release your insert, then use a compatible pair of enzymes to prepare a plasmid vector to receive the fragment. Now there's a way to do it automatically! If you infect male bacteria with both your lambda ZAP II clone and also a filamentous helper phage, the trans-acting proteins in the filamentous phage will cause replication of the central portion of the lambda clone, yielding a single-stranded DNA product that is packaged and secreted in a filamentous phage coat. The plaques therefore have TWO count 'em TWO types of phage:

1. The original lambda ZAP II clone

2. A single-stranded copy of the central region, in a filamentous phage capsid

How does this work? Having an f1 origin of replication in a vector makes it sensitive to the replication machinery of bacteriophage f1. You can cause packaging of single-stranded DNAs amplified from the f1 origin, even though there are no other elements of the phage present on the DNA strand.

What next?

You take the plaque containing these two types of phage and plate them on a bacterial strain that is resistant to phage lambda but sensitive to filamentous phage (i.e. use male bacteria), and don't forget to add ampicillin to the medium. The bacteria will take up the filamentous phage carrying your favorite DNA (the single-stranded central portion of the lambda ZAP II clone), convert it from a single-stranded to a double-stranded molecule, and treat it just like any other plasmid. Since this new plasmid contains the

ampicillin resistance gene, the cells maintaining it will survive on the selective medium. This is a case of phage transduction, but it does not lead to further generation of infectious particles. The introduced plasmid contains the f1 origin of replication but no other part of the filamentous phage. There are no longer helper phage around, so the sequence is not packaged and secreted as it was before. It is now simply a stable plasmid, that looks something like this:

pBluescript II KS plasmid vector

pBluescript II KS (+/-)
phagemid vector
2961 bp
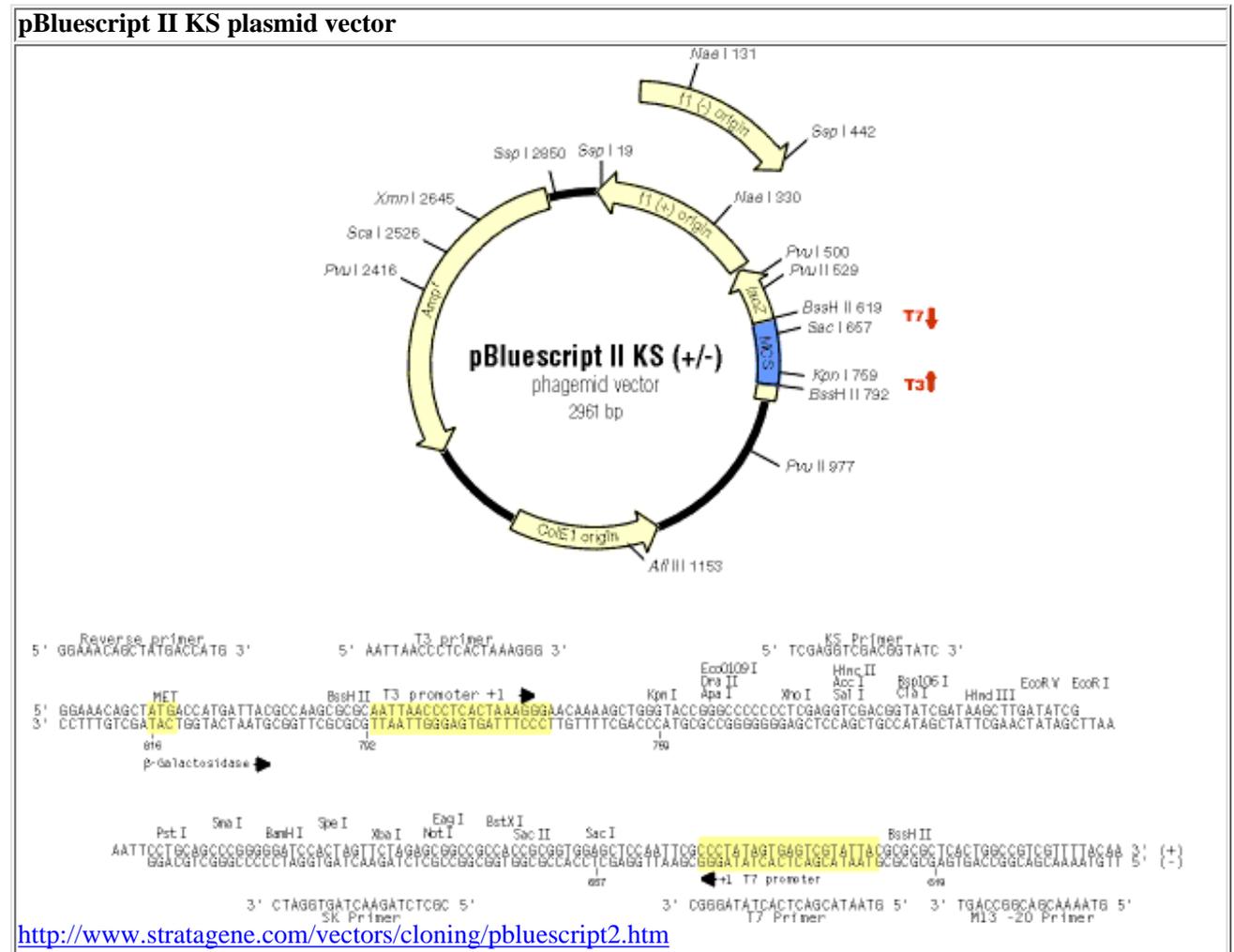
Is there a term that describes this interchange between double-stranded plasmid and packaged single-stranded filamentous phage particles? How about "complete and utter confusion"? Actually there's a better word than that. We call a plasmid that can be turned into a transducing phage (and vice versa) a "phagemid" to indicate that it is a sort of hybrid.

Plasmid vectors Plasmids are a bit more problematic, because lifting a membrane from a collection of colonies may strip away the entire colony, leaving you nothing to grow later. One solution is to make an imprint of the live colonies on a piece of autoclaved velvet, then print the colony pattern onto a reference plate, a process called "replica plating," but this is an unpleasant task. Plasmids vectors are generally not used for "expression systems" like gt11, and are typically used for DNA detection only (i.e. using DNA probes to detect a specific clone rather than antibodies). In the diagram below, the plate at left contains a collection of bacterial colonies that are first replicated to velvet and printed on a clean reference plate (starting with step 1, going horizontally to the right). After the plate is replicated, the remaining cells in the colony are adsorbed to a piece of

nitrocellulose (step 2), and the colonies are lysed directly onto the nitrocellulose (or nylon). This imprint can then be probed with a DNA probe, just as you might handle a Southern blot, and you may find that one of the colonies corresponds to a positive hybridization result.

velvet imprint   reference plate
Nitrocellulose lift   DNA probe results
correlation
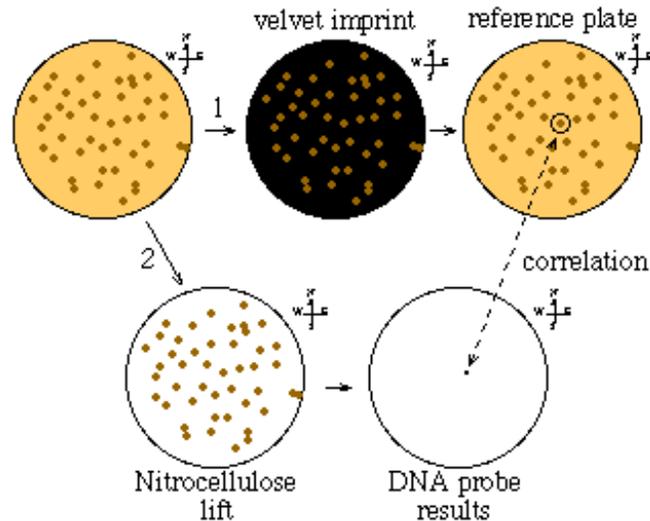
Lucky thing that you remembered to replicate the plate before adsorbing the colonies to nitrocellulose - now you have a reference colony you can pick and grow!

A second problem with standard laboratory plasmids is that it is difficult to clone large insertions into them, and this is particularly a problem if the library consists of genomic DNA fragments rather than cDNAs. Most plasmids do not grow well if they have large (10 kbp) insertions. While that is similar to the limit for gt10 and gt11, we'll be discussing later, lambda phage that can handle up to 23 kbp of DNA. The problem of "insertion size" and representation in the library may become more clear from the next example.

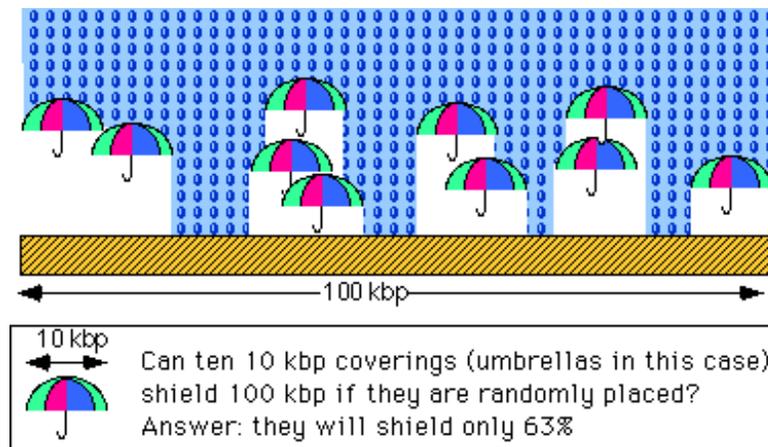*How many different recombinants must a library have?* Since a library contains random fragments, you can never be 100% certain that it contains a clone you desire.

Here's an example of the problem. Suppose you have a genome of 100 kbp, and you break it into random fragments of about 10 kbp for cloning (of course a real genome would be much larger). If you obtain a library with ten phage in it, would the entire genome be represented?
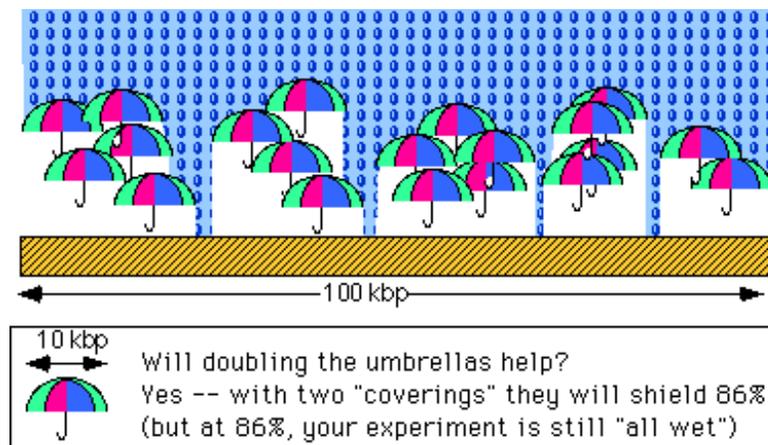
The answer is no! Some of the clones would overlap.

Let's try the same problem, substituting umbrellas for DNA clones:

Suppose we want to protect a group of movie stars from the rain, while they're walking from their limos to the doors of the Dorothy Chandler Pavillion (a journey of 100 killibradpitts (kbp) - a new measure of distance). We have 10 umbrellas on hand, each able to cover 1/10th of the distance from the street to the door. If we distribute them randomly, will they cover the entire path and keep the stars dry?

10 kbp
Can ten 10 kbp coverings (umbrellas in this case) shield 100 kbp if they are randomly placed?
Answer: they will shield only 63%

Of course, if we are thinking that umbrellas = clones and sidewalk = genome, then some sequences in the genome will not have been cloned, and will not be represented in your library. That's fine, so long as the sequences that you want to clone are not left out, but we are not always so fortunate. What would happen if we made the library twice as big? Let's go back to the Academy Awards, and see how we fare with twice the coverage:

10 kbp
Will doubling the umbrellas help?
Yes -- with two "coverings" they will shield 86%
(but at 86%, your experiment is still "all wet")

As you can see, we got closer to complete coverage, but still have some unprotected areas. To be precise, only 86 killibradpitts of sidewalk are covered and 14 are uncovered.

With our analogy of clones, the uncovered regions represent uncloned sequences in the genome. The way we can calculate what's missing is pretty easy -- it's based on the Poisson distribution. If we have one "covering" of the genome (as in the ten clone example) the chance of a sequence being cloned is (1- 1/exp(1)) or about 0.63 (or 63%), where exp(x) = the base of the natural logarithm e to the x power, . If we have two "coverings" of the genome (as in the twenty clone example) the chance of a sequence being cloned is (1- 1/exp(2)) or about 0.86 (or 86%), where exp(2) means the square of e. The pattern continues as your library contains more independent clones. With ten

"coverings", the chances reach 0.99995 (or 99.995%), but you can see that you never reach absolute certainty of cloning any particular sequence.

More on the walk of the stars next time!

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