

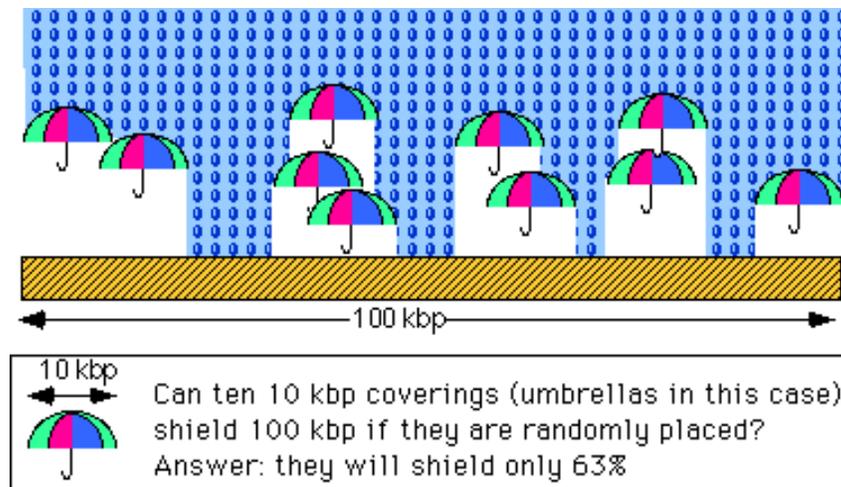
Lecture 17

Virulent lambda vectors

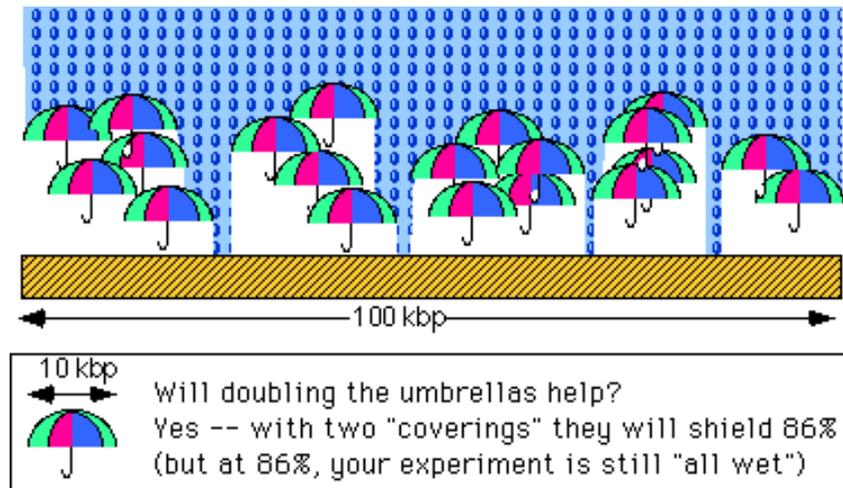
Size does matter

How many different recombinants must a library have? Let us carry on this discussion about the representation of libraries. In the last lecture I mentioned that since a library contains random fragments, you can never be 100% certain that it contains a clone you desire.

I told you a story in which umbrellas substitute for DNA clones, with the goal of protecting a group of movie stars from the rain. 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?



No, with only "one covering" on hand, they only keep $1 - (1/e)$ dry. If you are an ant on the big golden carpet, there is a 63% chance that you are being kept dry. If the number of umbrellas is doubled, the protection from the rain goes from 63% to 86%, as $1 - (1/e)^2$



Sounds pretty fishy to me, but that's the Poisson distribution.

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.

If I have a library with 100 phage in it, and I grow it in bacteria so that, following lysis, I have 10,000 phage, will I have a better chance of getting a clone I

No, your chances are not improved. The issue of representation in a library deals with a count of the independent phage clones. If you make copies of the original set, you still only have 100 independent phage, even if you "amplify" the library so your tube now has 10,000 phage.

Amplification of a library has certain risks, in that poorly growing phage become more rare in comparison with rapidly growing phage. As a result, you can't be certain that 100 phage from an amplified library represents the same fraction of the genome as 100 phage from an unamplified library.

Amplification is used to preserve a library so it can be used many times -- not to improve the likelihood of success in screening.

Here are a couple of problems to think about:

Suppose the library was based on an organism with genome size 3×10^7 bp and had been prepared in a plasmid vector, with insertions of 5 kbp. How many colonies would we have to plate to have 99% certainty of cloning a sequence from *T. brucei*?

Answer: Since each independent element of the library contains 5 kbp, which is $1/6000$ of the genome, one "covering" is 6000 clones. To achieve the figure of 99% certainty, we need 4.61 coverings (because $1 - 1/\exp(4.61) = 0.99$), so we need $4.61 \times 6,000 =$

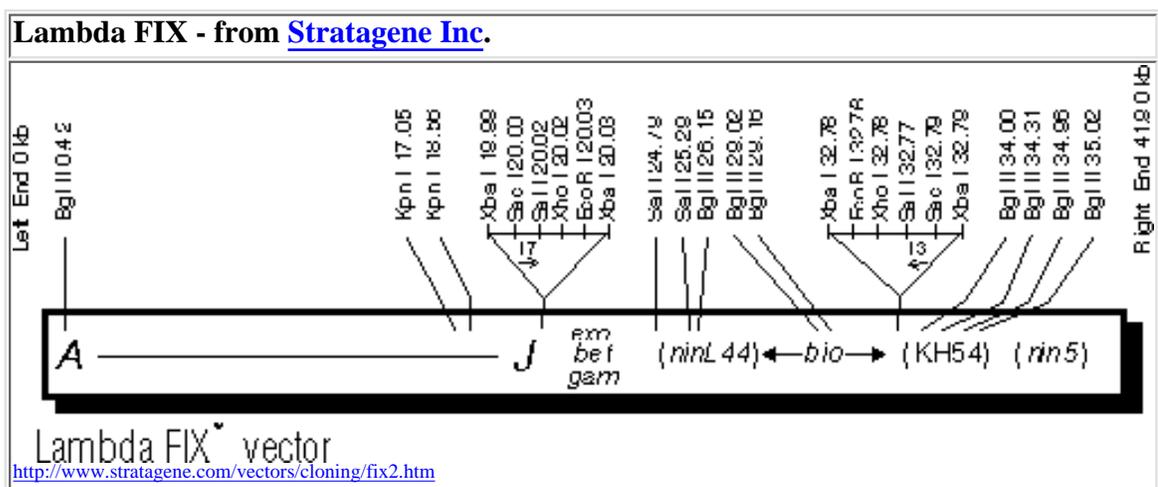
want?

27,660 independent colonies.

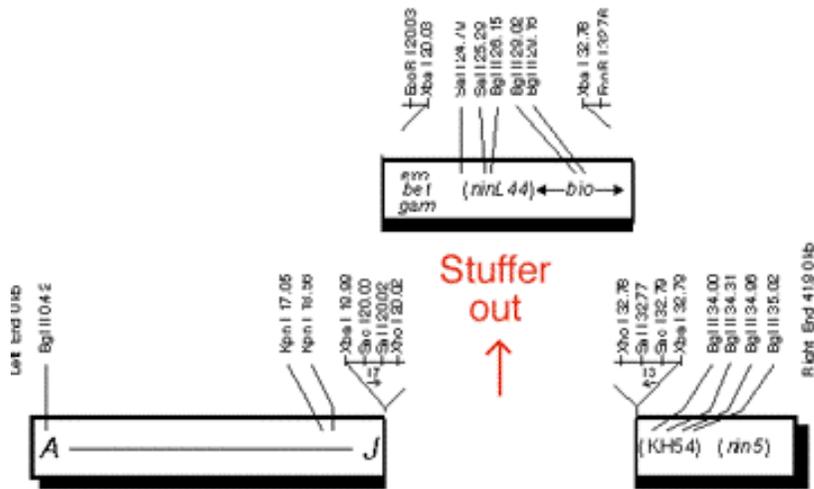
Suppose we have a phage library with random 20 kbp fragments of the human genome, which is 3,000 Mbp in size (3 billion base pairs). How many independent phage would we need to screen to have a 99% certainty of cloning any particular human sequence?

Answer? I'll leave this one to you to solve

Virulent lambda vectors We've already seen several advantages to having a temperate phage vector. In the case of lambda gt10, the ligation products lacking insertions can be selected or screened out by virtue of their lysogeny. In the case of lambda gt11 and ZAP vectors, maintaining a clone in a lysogenic state can minimize expression of a potentially toxic protein product from the inserted sequence. Sometimes, cloning pieces of DNA 9 kbp at a time is just too impractical. To clone bigger pieces (9 to 23 kbp) you need a "stripped down" version of the phage. For example, [take a look at lambda FIX](#).



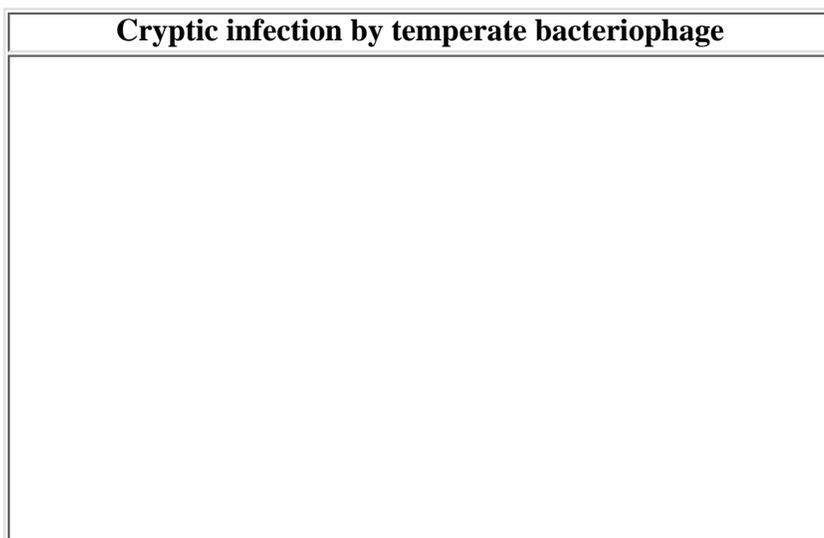
Note that there appear to be two polylinkers; one at 20.00 kbp and the other at 32.78 kbp. In fact, the sequence between the polylinkers (*ninL44*, *bio*, etc.) is a "stuffer" fragment that is discarded. The purpose of the stuffer fragment is just to serve as a "placeholder" while the vector is being replicated as a phage.

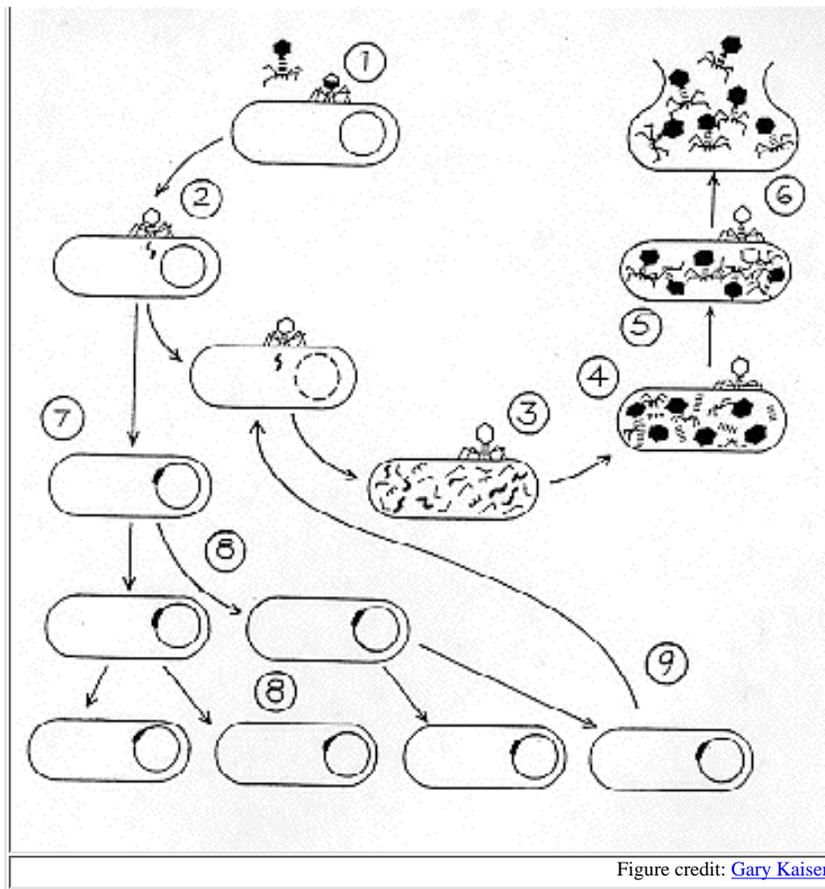


Don't forget that lambda phage are only viable if they contain between about 39 and 52 kbp of DNA. With the 14 kbp stuffer present, the FIX sequence would amount to 43 kbp which would make a viable phage. Without the stuffer, the remaining 29 kbp would be too small to make a viable phage! The stuffer is really like that little piece of white cardboard under the "Twinkee." The cardboard helps the product keep its shape, but when you're ready to eat, you throw it away.

Where did the *cI* gene go? It was left out! Everything not needed for virus production was eliminated from this vector, so that there would be extra room for foreign DNA inserts. Because of this consideration, the amount of lambda DNA in the two arms is 29 kbp, leaving up to 23 kbp free (because 29 kbp + 23 kbp = 52 kbp, which is the maximum size). What is the consequence of leaving out these sequences? The virus can only grow lytically, using steps 1-6 in the figure below.

Reminder: The split life cycle of wild type lambda

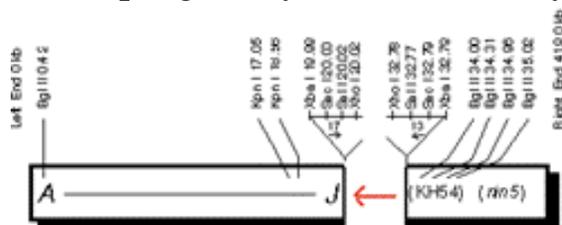




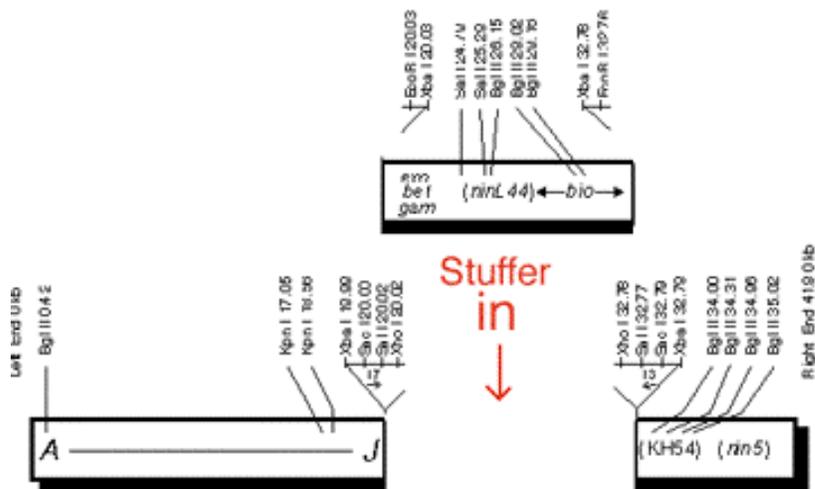
Since lambda FIX cannot use steps 7-9 of the life cycle, it is now purely virulent, and no longer temperate.

The cloning steps used with lambda FIX are exactly the same as described previously for lambda ZAP (and gt11, and gt10). You prepare fragments of DNA using one of the enzymes shown in the polylinker (or at least arrange to add appropriate linkers or adapters), ligate the fragment(s) into the prepared arms of the phage, and package the resulting concatemers into phage capsids.

Two legitimate concerns: What happens if the two phage arms ligate together without an insert between them? Nothing! The two arms combined are too small to be packaged into a viable phage, so you automatically select for phage with insertions.



What happens if the stuffer is recloned in the phage arms, instead of your added DNA fragment? This is certainly a problem, but there are several solutions.



1. If you are using Sac I or Xba I as a cloning enzyme, you can then also cut the phage with Sal I, which only digests in the stuffer region between those two restriction sites. Your stuffer DNA will be cut into small insignificant pieces with ends that aren't compatible with your cloning site, and will cease to be a problem!
2. You can separate the arms (10 and 20 kbp) from the stuffer (12 kbp) by gel electrophoresis in a low melting point agarose gel.
3. You can use the Xho-1 partial fill-in trick, provided that you are using Xho-1 as the cloning enzyme, and one of the compatible enzymes it will match (e.g. a partially filled-in BamHI site). Remember this method from a previous lecture?

The Xho I "partial fill-in" reaction	
Before digestion with Xho I	GAGGCTCGAGAATAC CTCCGAGCTCTTATG
After digestion with Xho I	GAGGC TCGAGAATAC CTCCGAGCT CTTATG
After partial fill-in with dCTP and dTTP	GAGGCTC TCGAGAATAC CTCCGAGCT CTCTTATG
5' GA overhang	Source
GATC NNNNNN AG NNNNNN	BamHI end, partially filled in with dGTP and dATP
GATCT NNNNNN AGANNNNN	Bgl II end, partially filled in with dGTP and dATP
GATC NNNNNN AGNNNNNN	Sau3AI end, partially filled in with dGTP and dATP
GATCA NNNNNN AGT NNNNNN	Bcl I end, partially filled in with dGTP and dATP

GATCYNNNNN	Xho II end, partially filled in with dGTP and dATP
AGRNNNNN	

The partially filled Xho-1 site is compatible with the partially filled sites shown above.

4. You can use an incredibly elegant trick, courtesy of your friendly neighborhood bacteriologist!

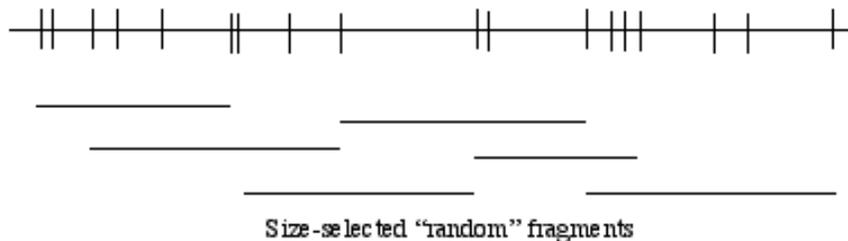
What trick might rescue us from the drudgery of screening clones with reinserted stuffers? Another sleight of hand with bacteriophage. It turns out that *E. coli* strains lysogenic for P2 bacteriophage (another temperate phage we haven't talked about) are unwilling to countenance having bacteriophage lambda in the same cell. It seems that P2 interferes with lambda infections because of the lambda red/gam genes. Lambda that are red/gam+ are **S**ensitive to **P**2 **I**nterference (said to have the **S**pi+ phenotype). Guess what? The red/gam genes are present in the stuffer fragment of lambda FIX! That means that lambda FIX carrying a re-inserted stuffer (i.e. those we don't want in our collection of clones) will be Spi+, and will therefore not grow on a P2 lysogenic strain of *E. coli*. Those lambda FIX carrying foreign DNA instead of stuffer will be Spi-, and will grow on a P2 lysogenic strain of *E. coli*. Very cool!

How do we get fragments for insertion?

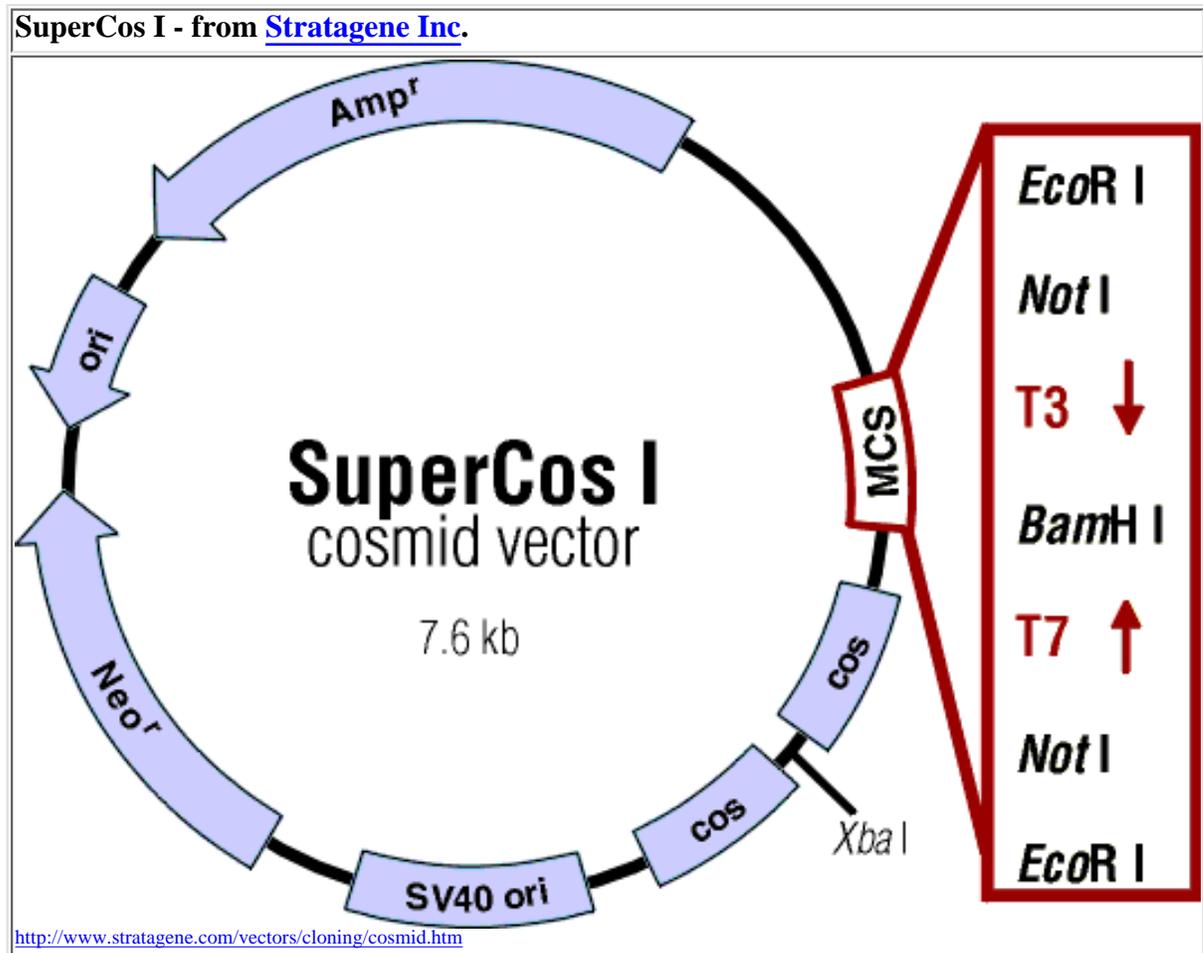
If you recall from our discussion, we want (ideally) to represent a genome with random overlapping fragments. Partial digestion with Sau3A1 is a good way to accomplish that, because Sau3A1 is a "four cutter" that digests the DNA frequently. It cuts at ^GATC

By adding a small amount of enzyme, not every site will be used (randomly) and we can generate a size selected collection of overlapping DNAs:

Sau 3A1 sites



More space! You can never have enough room to clone your favorite piece of DNA, it would seem! What prevents us from simply taking over all of the space lambda could offer in its viral capsid? If we could just fill the capsid with a big plasmid having cos ends (necessary for packaging) then we would have about 42 kbp of free space instead of only 23 kbp (as in lambda FIX). In fact, that kind of cloning vector has been made already, and it is called a "**cosmid**" (where the "cos" indicates that it has lambda cos ends). Here's an example of a commercial vector based on cosmid technology:



Let us look at the anatomy of this vector:

1. A sequence marked "ori" for DNA replication in bacteria
2. Amp^r for ampicillin selection in bacteria
3. A sequence marked MCS (multiple cloning site) that is a polylinker containing unique restriction sites and two phage RNA polymerase promoters (T7 and T3) in opposing directions.

4. Two cos sequences, separated by an Xba1 site
5. An origin of DNA replication from SV40 (permits replication and copy number amplification in many eukaryotic cells, in the presence of SV40 T antigen protein)
6. Neo^r for selection in eukaryotic cells with the neomycin antibiotic analog G418.

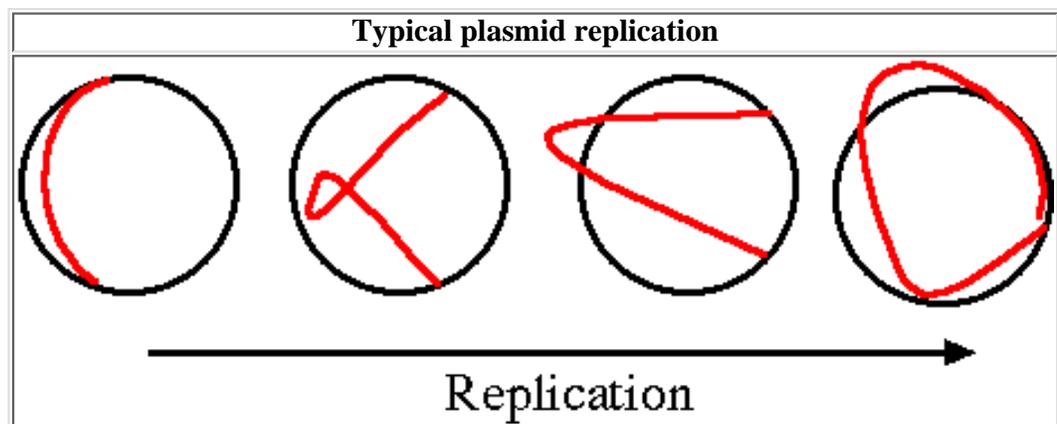
Numbers 1-3 are what we discussed at the beginning of the course, as being necessary elements. Number 4 permits packaging of the plasmid into lambda phage capsids in an in vitro packaging system. Numbers 5 and 6 are analogous to 1 and 2, but work in eukaryotic cells (some of the time at least). This is a new concept for us - a plasmid that works in two different types of cells. We call this a **shuttle vector**, because it can shuttle back and forth between the hosts. Of course, we don't use bacterial transformation or phage transduction methods to introduce DNA into eukaryotic cells. We'll discuss the differences later in the course.

The maximum amount of DNA that can be inserted into SuperCos I depends on the packaging limit of lambda (52 kbp) and the pre-existing size of the vector (7.6 kbp).

How much space does SuperCos I have for foreign DNA?
--

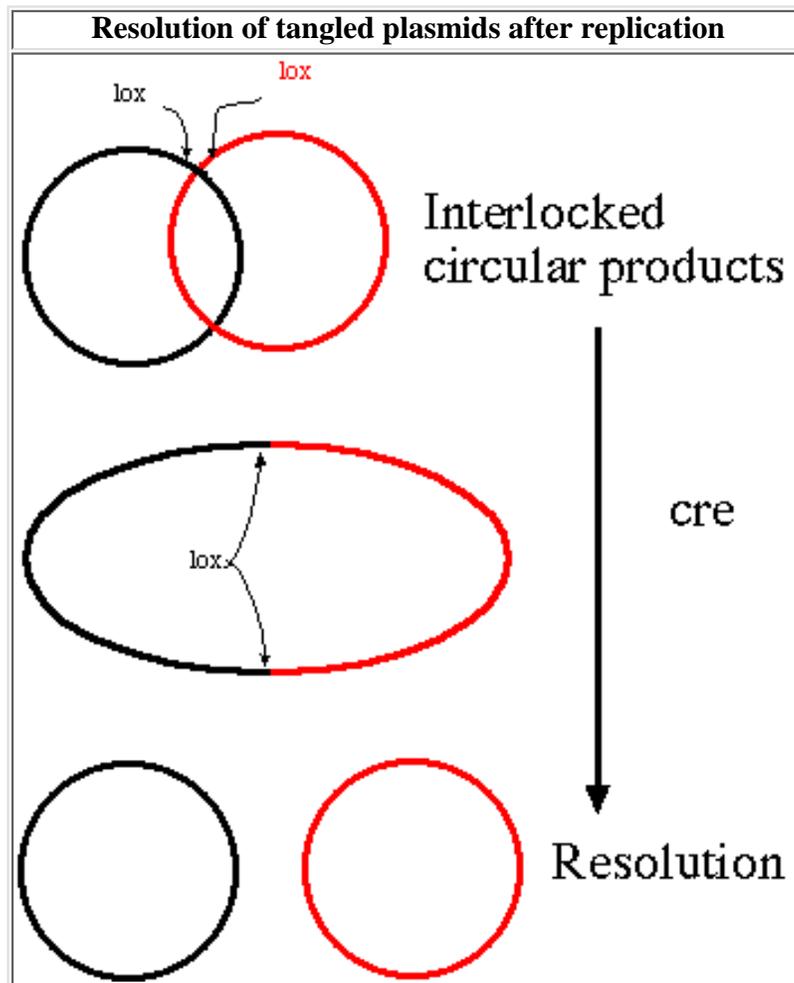
$52 \text{ kbp} - 7.6 \text{ kbp} = 44.4 \text{ kbp}$

P1-based If we want to package more than about 45 kbp, we'll have to turn to a phage other than lambda! Bacteriophage P1 is now commonly used as a cloning vector, and has several interesting features. The phage lysogenizes bacteria (it is temperate) but it doesn't usually integrate its DNA into the E. coli genome. Instead, it maintains itself as a plasmid. When it enters a lytic cycle, it makes many copies of its genome which have to be disentangled after replication:



This results in tangled plasmids that need to be resolved into individual unlinked monomers. The protein **cre**, which is a gene product of phage P1, finds "**lox**" sites in each of the tangled genomes and causes recombination.

After two rounds, the plasmids are untangled.



With that knowledge, let's study the structure of a P1-based vector from RPCI in Buffalo. These vectors can take large insertions of approximately 80 kbp.

Here's an example of a vector for P1 packaging:

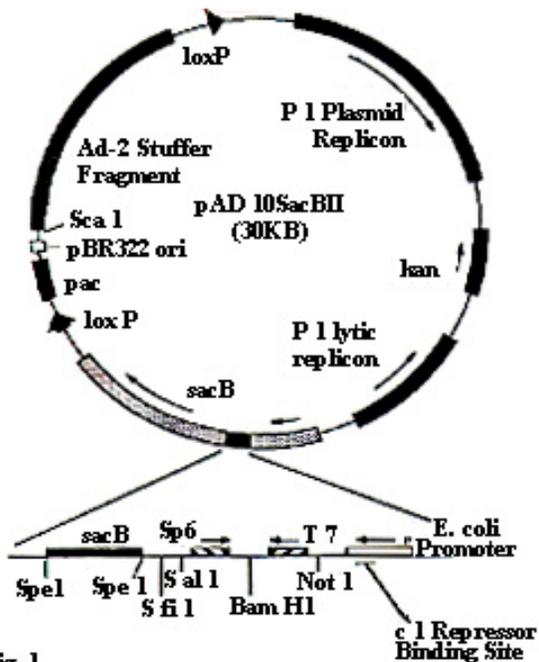


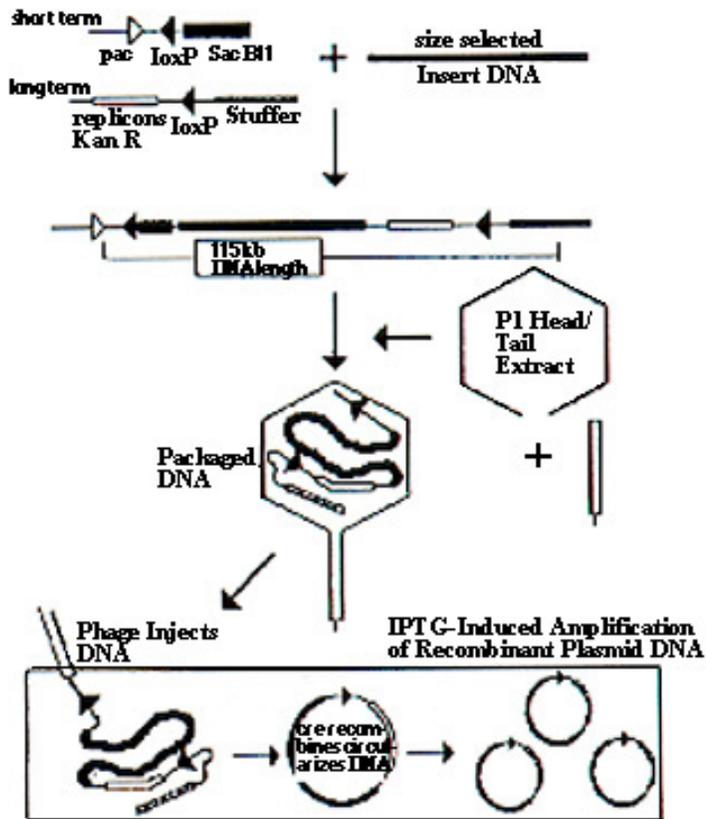
Fig 1 Plasmid map of pAd10-SacB II and cloning region

source: http://www.nalusda.gov/pgdic/Probe/v1n3_4/clon1.jpg

It is important to point out some of the features of this vector:

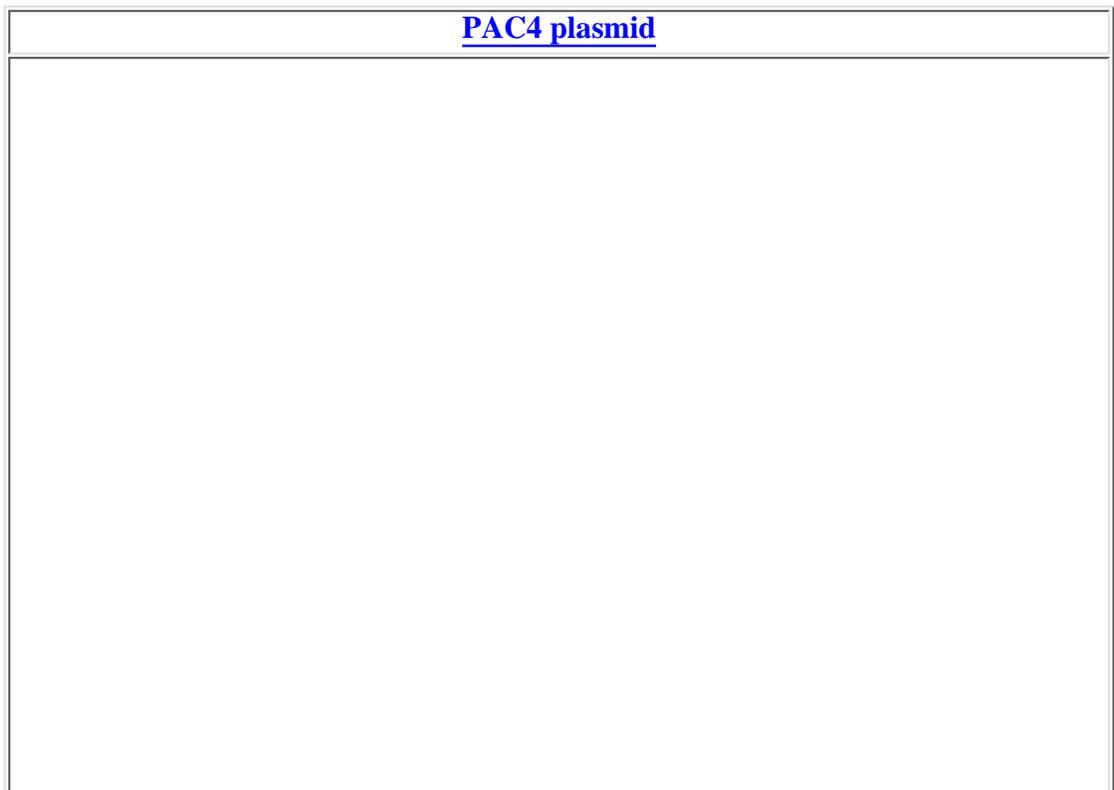
- PAC - a site for initiation of packaging in a P1 capsid (in vitro)
- Cloning site
- P1 lytic replicon
- Kan gene (for resistance in E. coli)
- P1 Plasmid replicon
- lox sites and stuffer from Adenovirus

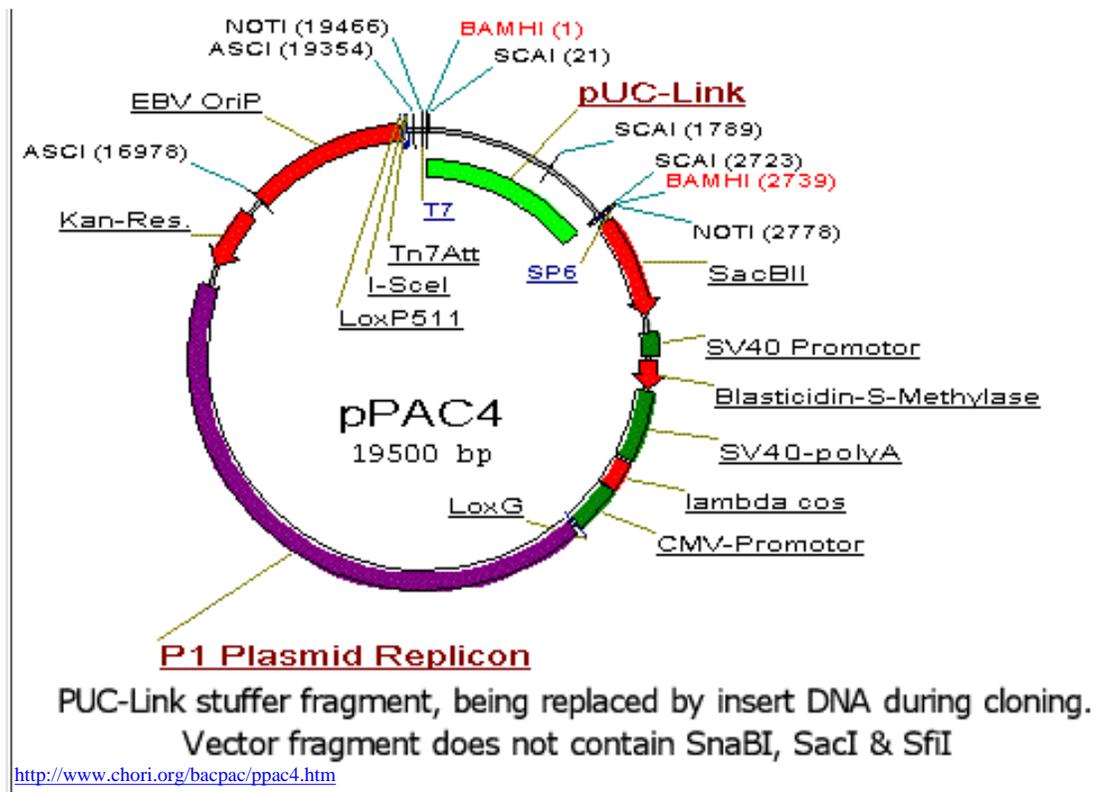
Here is a flow-chart showing how this can be used:



source: http://www.nalusda.gov/pgdic/Probe/v1n3_4/clon2.jpg

A P1 packaging system can be used for transduction of P1 phage-based plasmids, however there are also P1 artificial chromosomes (called **PACs**) that are larger than the packaging limit and are simply based on the replicon of P1. These may be handled by transformation.





Vector pPAC4 Information

Notes from: [RPCI](#)

The pPAC4 vector was constructed by Eirik Frengen at Roswell Park Cancer Institute, in the laboratory of Pieter J. de Jong. The vector contains a number of potentially very useful elements.

1. The low copy number replicon of bacteriophage P1 ("plasmid replicon"), which can stably propagate most foreign DNA sequences in E.coli.
2. The kanamycin resistance gene (only with E.coli promoter, thus no G418 resistance).
3. The BamHI cloning site (insert sequences will replace the small 2.7 kb BamHI fragment).
4. Flanking T7 and SP6 promoter sequences.
5. Two NotI restriction sites immediately flanking the T7-BamHI-SP6 segment.
6. The yeast intron-encoded nuclease recognition site (PI-SceI), which has no cut sites in the human genome, thus permitting linearization of all recombination clones irrespective of insert size or content (useful for optical mapping procedures).
7. The Epstein Barr Virus replicon (EBV oriP), for replication in human cells (require presence of EBV EBNA-1 protein to be provided in trans).
8. The 34 bp site-specific recombination site loxG (recombines with wildtype loxP sites catalyzed by the cre recombinase, and has a CMV promoter + ATG start codon in the vicinity). In vivo site-specific recombination with a complementary lox site in the mammalian genome (coupled to truncated, promoter-less neo gene) allows activation of G418 resistance under control of the CMV promoter and ATG startcodon.
9. The mutated loxP511 site (recombination proficient relative to similar loxP511 sites but less proficient with respect to the wildtype loxP site).
10. The blasticidin-S-methylase gene under SV40 promoter control, providing resistance to blasticidin in mammalian cell culture. The file ppac4.txt has a GenBank compatible file with the putative sequence of the pPAC4 vector (partly determined by sequencing various new cloning junctions and new elements, and partly based on assumptions on sequences from other elements already present in GenBank).

I'll attempt to decode their technical notations:

1. The origin of replication that is functional in *E. coli* cells comes from P1 phage (the P1 replicon). Unlike the ColE1 origin, the P1 origin generates a low copy number of plasmids per cell.
2. Kan-res stands for the kanamycin antibiotic resistance gene, for selection of maintenance in *E. coli*.
3. A pair of BamHI sites surrounding a "stuffer" fragment.
4. Flanking phage RNA polymerase sites, as we've discussed.
5. Two rare 8-cutter sites (Not I) flanking the cloning site.
6. An extremely rare Sca I site (from the yeast mating type locus restriction endonuclease). This site is so rare, it doesn't cut even once in the human genome. That's terrific if you want to linearize a library of random inserts, and don't want to risk digesting your inserts internally.
7. An origin of replication that is functional in (some) eukaryotic cells.
8. A lox site that can be used for integrating DNAs into a genome in a site-specific manner.
9. A drug-resistance gene that is suitable for eukaryotic cells.
10. Note a few more elements: an SV40 polyadenylation site, a transcriptional promoter from cytomegalovirus (CMV), and lambda cos ends.

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

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