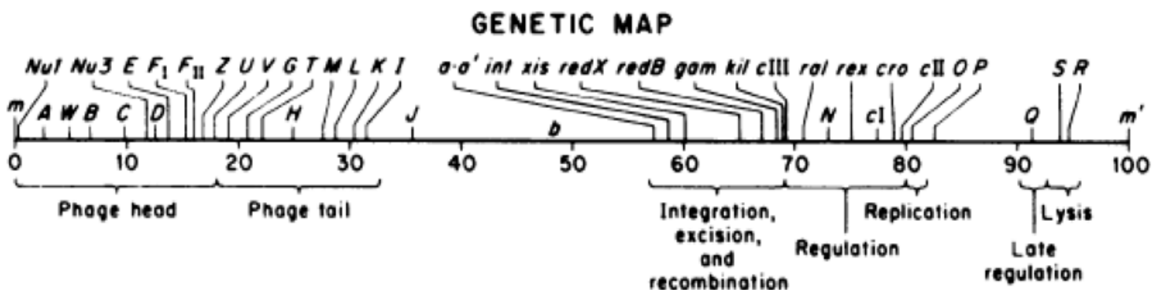
A short primer on cDNA libraries, bacteriophage lambda vectors and phagemids

During this course, we will be using a cDNA library that was constructed in the "HybriZAP" vector engineered and sold by Stratagene Inc. (www.stratagene.com). This vector has a number of unique properties that make it extremely useful for constructing and screening DNA libraries that will be used in the yeast two-hybrid system. First, it functions as a classic bacteriophage lambda cloning vector that enables easy cloning of large numbers of inserts and the formation of large representative cDNA libraries (see below). Second, it uses some "tricks" of bacterial and phage genetics to enable the excision of a hybrid "phagemid" vector out of the original lambda vector. This "phagemid" vector can replicate and function via two pathways: first, it can replicate and be packaged as an "f1" or "M13"-type bacteriophage (this is the "excision" part of phagemid excision); second, it can then be forced to replicate and be used as a normal plasmid vector. Finally, this plasmid vector can be easily tranferred to and replicate in yeast cells.

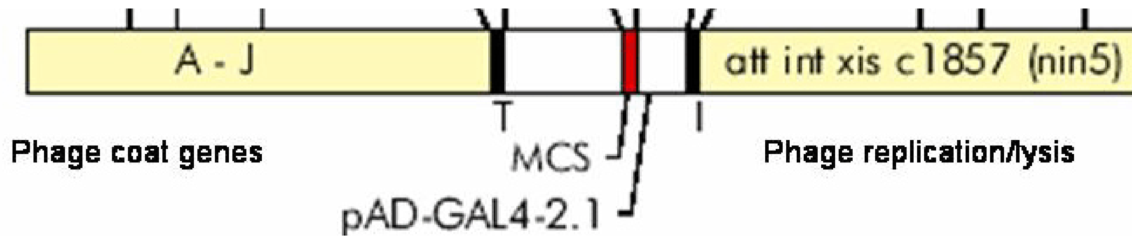
A. Bacteriophage Lambda

Bacteriophage lambda is a ~50 kb (50,000 bp) bacteriophage that infects *E. coli*. Wild type bacteriophage lambda (or just "lambda" = λ) has two alternative life cycles. In the "lytic" cycle, lambda infects *E. coli*, replicates λ DNA, translates phage proteins and lyses the bacterial cell, releasing new bacteriophage particles that can then infect additional bacterial cells. This mode of replication produces bacteriophage plaques (cleared areas) on a lawn of bacteria, as we saw when we titered the bacteriophage lambda library. The other life cycle is known as "lysogeny". In this life cycle, the λ genome inserts into the *E. coli* chromosome via recombination, and is replicated as a silent lysogen without damaging the host cell. As can be seen on the genetic map below, the λ genome can be divided into three sections. The "left arm" (map units 0 to 40 on the diagram) encodes protein subunits of the phage head and tail. The right arm (map units 70 to 100 on the diagram) encodes proteins and regulatory sites needed for phage replication. The middle section (map units 40-70) encodes genes required for lysogeny.



Because the central region of the lambda genome encodes functions dispensable for the lytic cycle (which is all we care about for using it as a vector), this region has been replaced in the HybriZAP vector by sequences that comprise the "pAD-GAL4-2.1" phagemid.

Diagram of HybriZAP Bacteriophage lambda vector



The above vector can be used as a standard bacteriophage lambda vector for cloning. For making a cDNA library cloned into the vector, the following steps are carried out:

1. mRNA is isolated from the tissue of interest (in this case mature anthers/pollen from *Petunia hybrida*
2. mRNA is copied to double-stranded cDNA using reverse transcriptase
3. "Linker" sequences (partially double-stranded oligonucleotides incorporating recognition sites for restriction enzymes) are ligated to the ends of the cDNA and the cDNA is then ligated (covalently inserted) into the multiple cloning site (MCS) of the vector
4. The ligated DNA is then packaged into bacteriophage lambda particles and used to infect *E. coli.*
5. The library that we started with consists of a collection of bacteriophage lambda particles, each of which contains a single cloned insert in the HybriZAP vector. The original size of the library was approximately 2 x 10^6 clones.

B. Library Size and Screening Considerations

One question that needs to be answered when constructing and screening DNA libraries (genomic or cDNA) is "how big does the library need to be"? That is, how many individual library clones does one need to have in order to have a reasonable assurance that (e.g. for the case of a cDNA library) that every single mRNA in a particular tissue is represented in the library? This is not a trivial issue, since many mRNAs, including some of the most interesting mRNAs, are present at quite low concentrations inside the cell. One can estimate the number of clones needed using an equation known as the "Clarke-Carbon" equation. This equation has the following form:

$$N = \frac{\ln(1 - P)}{\ln(1 - f)}$$

where N= the number of clones needed, P = probability of having the desired clone (usually set to 0.95 to 0.99), and f= the fraction of the sample made of of any single clone. This last is usually the most difficult number to estimate. Some examples illustrate the point best. For a cDNA library, we can estimate from earlier nucleic acid reassociation data, that a typical cell contains approximately 100,000 individual mRNA molecules. Thus, "f" for a mRNA present at 1 copy per cell would be 1/1 x 10^5 or 1 x 10^{-5}. A typical genome from a higher eukaryote contains

approximately 1.5 x 10^9 bp of DNA. If the average insert size for a genomic clone is 15,000 bp, then "f" for a single clone in a genomic library would be 15000/1.5 x 10^9 = 1 x 10^{-5}. Using the above numbers, we can carry out some sample calculations:

1. Assume that we want to make a cDNA library that has a 99% probability of having every mRNA in the cell represented. How many clones do we need at a minimum? For this calculation, P = 0.99 and f = 10^{-5} (see above). Thus the above equation is:

   $N = \ln(1-0.99)/\ln(1-1 \times 10^{-5})$ and $N = 4.6 \times 10^5$ or about 460,000 clones

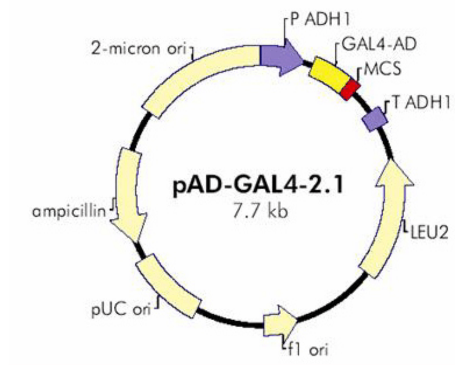
2. Assume that we want to make a genomic DNA library. The genome size of our organism is 1.8 x 10^9 bp and the average insert size in our vector is 18 kb. How many clones do we need to have a 95% probability of cloning every gene in the genome? We would use the values P= 0.95 and f = 1.8 x 10^4/1.8 x 10^9 = 1 x 10^{-5}. Thus, N = 3 x 10^5 or about 300,000 clones

These are relatively large numbers, but it is fairly simple to make libraries of this size when using bacteriophage lambda vectors (because of the ability to "package" large numbers of bacteriophage particles *in vitro* and then use those particles to infect bacteria. It is much harder to make libraries of this size using conventional plasmid vectors that have to be chemically transformed into *E. coli.*

C. Phagemid (phage/plasmid) Vectors

So, we can see the utility of using bacteriophage lambda vectors for making libraries, since we can fairly easily obtain lots of clones, and thus have large, representational libraries. The problem is, however, that any indvidual clone from a lambda library is difficult to work with. First it's big (50 kb total), second, the amount of lambda DNA that can be isolated from a bacterial culture infected with phage λ is fairly small, and third (especially for a cDNA library) the ratio of insert size to total size of the vector can be quite low (e.g. 1000 bp insert - a typical mRNA - in a 50,000 bp vector). Routine cloning manipulations are much easier in plasmid vectors, which are often as small as ~ 3000 bp in size, but as we have discussed are poor choices for making large libraries.

The solution? Make a "hybrid" vector. The HybriZAP vector used to make the cDNA library we are using functions initially as a typical bacteriophage lambda vector (the original size of the cDNA library that was constructed was about 2 million individual clones), but then can be induced to excise a plasmid vector pAD-GAL4-2.1 (shown below).



The above vector has properties that make it (a) excisable from its bacteriophage lambda "parent" vector, HybriZAP, (b) able to replicate in E. coli for routine cloning and manipulation, and (c) able to replicate and transcribe proteins in yeast. (We will discuss the details of how this vector functions in yeast at a later time.) Key sites on the above vector that give it the properties listed are:

1) The f1 bacteriophage origin of replication (f1 ori)
   When co-infected with a "helper" phage (ExAssist) that provides the needed replication proteins, this vector is replicated from the f1 bacteriophage origin of replication, and is packaged into f1 bacteriophage particles. These bacteriophage do not lyse the bacterial cell, but are extruded out of the cell into the media.

2) An origin of replication that functions for standard plasmid replication in *E. coli* (pUC ori).

3) An origin of replication that functions for plasmid replication in yeast (2-micron ori)

4) An antibiotic resistance gene for selection in bacteria (ampicillin resistant)

5) An auxotrophic gene (LEU2) for selection in yeast (cells with this plasmid can grow on media lacking the amino acid leucine).

D. <u>Phagemid Excision</u>

The steps of going from a bacteriophage lambda vector to a plasmid vector require excision of the "phagemid" vector pAD-GAL4-2.1 from the HybriZAP vector. This excision occurs in two stages and depends on functions provided by helper phage and on specific genes that are present in "permissive" host cells (XL-1 Blue MRF') and missing from non-permissive host cells (XLOLR).

1. Stage 1 - excision of phagemid and packaging of phagemid particles
   A culture of permissive host cells (XL-1 Blue MRF') is infected with a mixture of bacteriophage lambda (library phage with HybriZAP vector) and helper phage (ExAssist helper phage). By mixing in a ratio of 1 (λ phage): 10 (XL-1 blue): 100 (ExAssist) we ensure that every lambda phage infects a bacterium, that no bacterium is infected by more than one lambda phage, and that every infected bacterium is also co-infected by helper phage. The helper phage drives replication of the pAD-GAL4-2.1 phagemid from the f1 origin of replication, via rolling-circle replication. This molecular replicates as a circular double-stranded molecule inside the bacterium and is packaged as a single-stranded circular DNA in a bacteriophage particle that is extruded out of the host cell into the media. The ExAssist helper phage contains an amber (nonsense) mutation. The protein with this mutation can be translated in XL-1 Blue MRF' because of the *supE44* mutation in this host that provides a suppressor tRNA that can utilize the amber codon as a regular amino acid codon.

2. Stage 2 - conversion to replicating plasmid
   The pAD-GAL4-2.1 vector does not encode any replication proteins. Therefore, it can only be replicated when DNA polymerases and initiator (primase) proteins are present from some other source - either the helper phage or the *E. coli* host. As discussed above,

when helper phage are present in excess, DNA replication occurs from the f1 origin of replication and the vector replicates and is packaged as an f1 bacteriophage. To convert the f1 phagemid to a replicating plasmid, phagemid particles (from the supernatant of the culture in step 1, above) are used to infect the non-permissive host XLOLR. This host lacks the suppressor tRNA that is present in XL-1 blue. Thus, replication proteins from the ExAssist helper phage cannot be translated, and the f1 origin of replication is not used. Instead, the vector is replicated as a standard plasmid vector in *E. coli*, using *E. coli* DNA polymerase acting at the pUC orgin of replication.