Bioinformatics and Bioengineering Summer Institute (2005)

Phage sequences in bacterial genomes: Spreading the tools of death

I. The Scenario: Construct a set of primers to detect recombinant Shiga toxin phages carrying variant tail fiber genes

Beneficial Bacteria vs. Pathogenic Bacteria

Some of our most potent tormenters come from among the enterobacteria, (e.g. Enteropathogenic *Escherichia coli* species (urinary tract infections; diarrheal disease, hemolytic uremic syndrome), *Yersinia pestis* (plague), *Salmonella enterica* species (typhoid fever; food poisoning), and *Shigella dysenteriae* (dysentery). Yet we live daily with trillions of harmless, even beneficial enterobacteria in our intestines. The line between the dangerous pathogens and their disarmed relatives is small. This is the case for bacteria producing phage-encoded toxins, such as cholera toxin and diphtheria toxin. The only difference between the benign and virulent bacterial strains in these cases is acquisition of a bacteriophage carrying the toxin genes (see Table1). If the critical determinant of pathogenesis is carried by a bacteriophage, then the spread of this phage can create new and potent pathogens.

Table 1: Known examples of bacteria with phage-encoded toxins

Corynebacterium diphtheriae	diptheria toxin
Clostridium botulinum	botulinum neurotoxin
Streptococcus pyogenes	erythrogenic toxin A
Vibrio cholerae	cholera toxin
Escherichia coli O157:H7	Shiga toxins

II. Properties of bacteriophages

What determines whether a particular phage can infect and introduce toxin genes into a certain bacterial host? An examination of the particular properties of the phage can shed light on what enables it to act as a toxin delivery vehicle. However, before one can ask how Shiga toxin phages recognize their host bacteria, it is useful to consider which properties phages have in common and which distinguish one from another.

II.A. General properties of phages

All bacteriophages (viruses that infect bacteria; also known simply as phages) are extremely small, well below the resolution of the light microscope. None are capable of growing outside cell, so they all have solved key problems: (1) How to travel from one cell to the next, and (2) how to influence cell behavior so as to maximize viral replication.

All phages carry nucleic acid, which may be RNA or DNA, double-stranded or single-stranded. The size of phage genomes varies from a bare four genes to 200 genes, not far from the number carried by tiny bacteria. The Shiga toxin phages belong to the most abundant class of phages:

double-stranded DNA viruses. Shiga toxin phages are related to the well studied bacteriophage lambda (and are therefore referred to as “lambdoid” phages), falling in the middle range with about 80 identified genes.

All phages must replicate their nucleic acid. A phage with linear DNA must solve the problem of how to replicate the ends. Eukaryotes solve this problem with telomeres. Phages have found other solutions. Lambdoid phages solve the problem by injecting their linear DNA with cohesive ends that enable the DNA to form an endless circle.

All phages must package their nucleic acid into protective heads (also called “capsids”, or “coats”). These serve both to maintain the nucleic acid in a compact state, reducing breakage, and to protect its genes from the environment. The capsid proteins are amongst the most conserved in the phage. Most phage (unlike animal and plant viruses) also have tails, through which to inject their nucleic acid into new bacterial hosts.

A newly recognized characteristic feature of the genomes of double stranded tailed bacteriophages, revealed by the recent availability of large numbers of phage genomes, is that they are highly mosaic. This means that different segments of the genomes match between different phages to different degrees (Fig 1). Thus, for example, the lambdoid phages λ and HK97 match each other at >95 % DNA sequence identity in the area of the *cI* and *cro* genes, but on either side of this region there is a sharp transition to a segment in which the two genome sequences do not match detectably at all. The mosaic boundaries that are seen in pairwise comparisons of genomes are taken to be the sites of illegitimate (non-homologous) recombination in the ancestry of one of the phages being compared (since a homologous recombination event would not give rise to a sharp transition in degree of similarity. The evident result of these recombination events is that individual phages of this family are genetic mosaics, drawing sequences from a shared pool. Regions of sequence similarity between phages provide further opportunity for genetic mixing via homologous recombination.

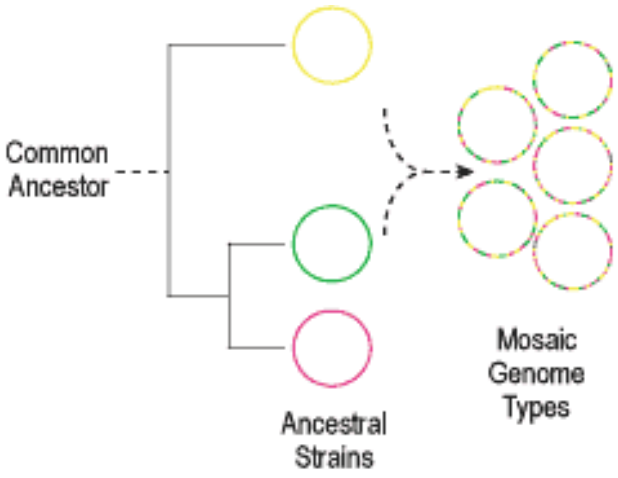


Fig. 1. An illustration of genetic mosaicism

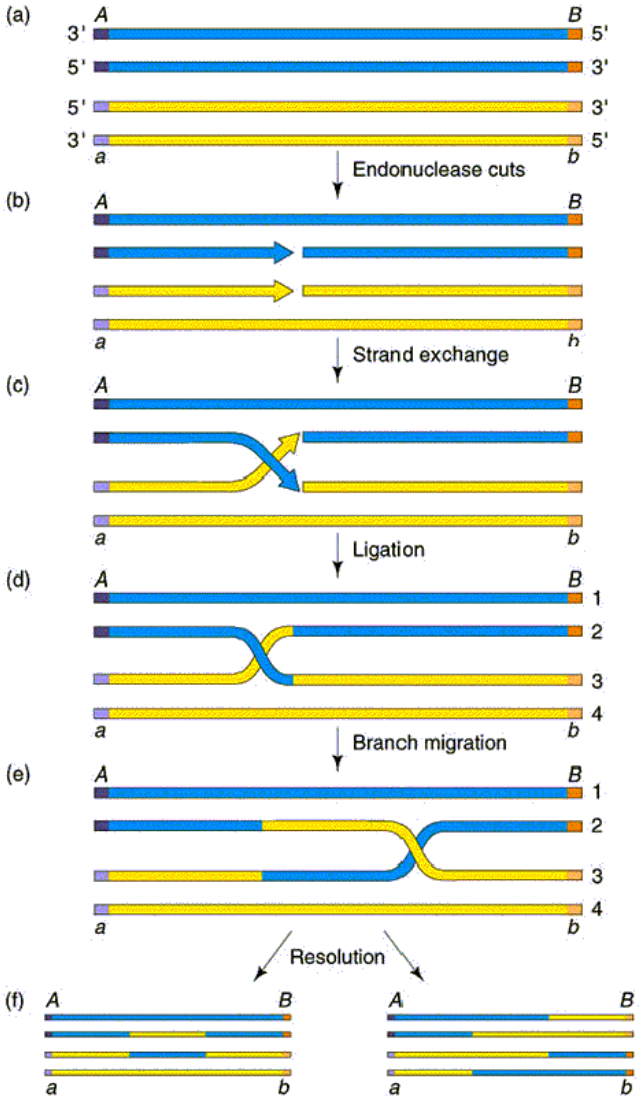


Fig. 2. Model for homologous recombination

Homologous recombination involves the pairing and exchange of identical (or nearly identical) sequences of DNA, resulting in hybrid progeny molecules (Fig 2, left). Illegitimate recombination occurs between non-identical (or non-homologous) sequences. In some cases it is thought that nearby homologous DNA may "anchor" the two molecules and allow the exchange to occur (Fig 3, below).

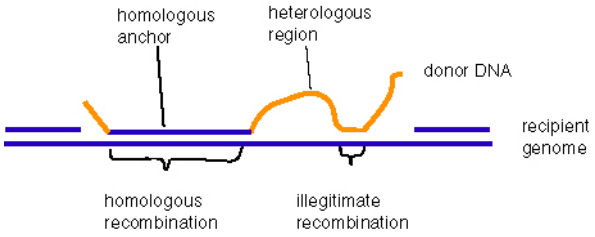


Fig. 3. Model for illegitimate recombination. In this example homologous recombination within the anchor sequence couples the donor and recipient DNA. This strongly increases the probability for an illegitimate recombination event that leads to the integration of a heterologous DNA segment.

II.B. Specific properties of phages

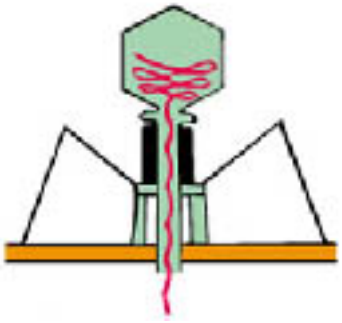
A phage is able to infect only a circumscribed set of bacteria, called its ***host range***. The host range is determined primarily by the surface receptor to which the phage binds prior to injecting its nucleic acid (Table 2). Some phage, like P1, adsorb to its hosts' lipopolysaccharide, a relatively well conserved component amongst bacteria; thus P1 has a relatively broad host range. Other phage, like phage lambda, adsorb to receptors that may be absent even in relatively close relatives of its host. Lambda adsorbs to a maltose transport protein and has a relatively confined host range. The host range of a phage limits the bacteria into which it can transfer genes.

Table 2. Examples of phage host range determinants

Phage	Receptor
T4, P1	LPS core polysaccharide
T2	OmpF porin protein
T1, T5	TonA ferrichrome transport protein
T6	Tsx nucleoside transport protein
lambda	LamB maltose transport protein
chi	flagellum (Salmonella)
f1, MS2, M13	F-pilus
SP-50	Teichoic acid (Bacillus, gram positive)

The phage binds to the host cell receptor sites through docking proteins at the end of the tail fibers. In some cases, additional specific contacts are then made through secondary receptors. Host cell specificity is conferred by the tail fibers of the bacteriophage.

Once firmly attached to the cell, a viral enzyme in the tail punches a hole in the host's cell wall (and for some phages the core region of the tail is thrust through) and the DNA is ejected into the host's cytoplasm (Fig 4).



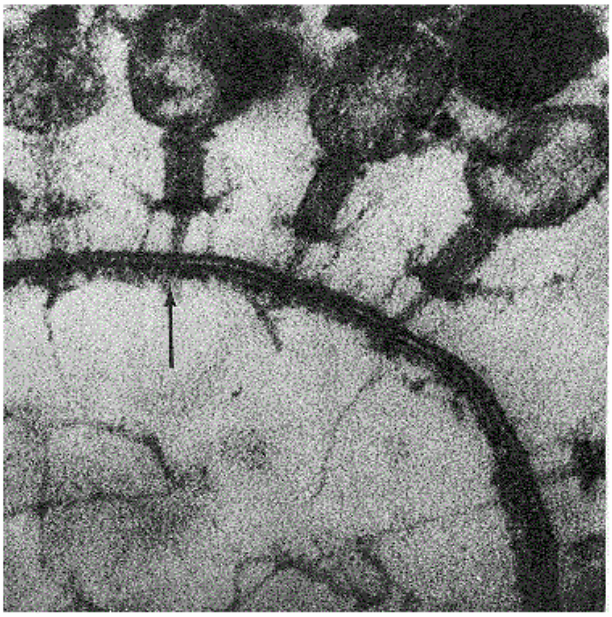


Fig. 4. Attachment and DNA injection by a tailed bacteriophage. Schematic (above), transmission EM of phage T4 adsorbed to *E. coli* (right).

Some phages inject their nucleic acid, replicate themselves, kill their host, and that's that. Others, including the lambdoid phages, live continuously with choice: to lyse (kill the host) or to lysogenize (coexist as a prophage)? In lysogeny (Figure 5), the phage forbears lysing the host and instead propagates itself within the cell through many generations of cell division. Lysogeny may offer the phage distinct advantages. Perhaps bacterial hosts are scarce, and killing the current host might leave the progeny phage with nowhere to go. Furthermore, tying its fate to the fate of the bacterium may be a winning decision if the bacteria is successful and rapidly growing.

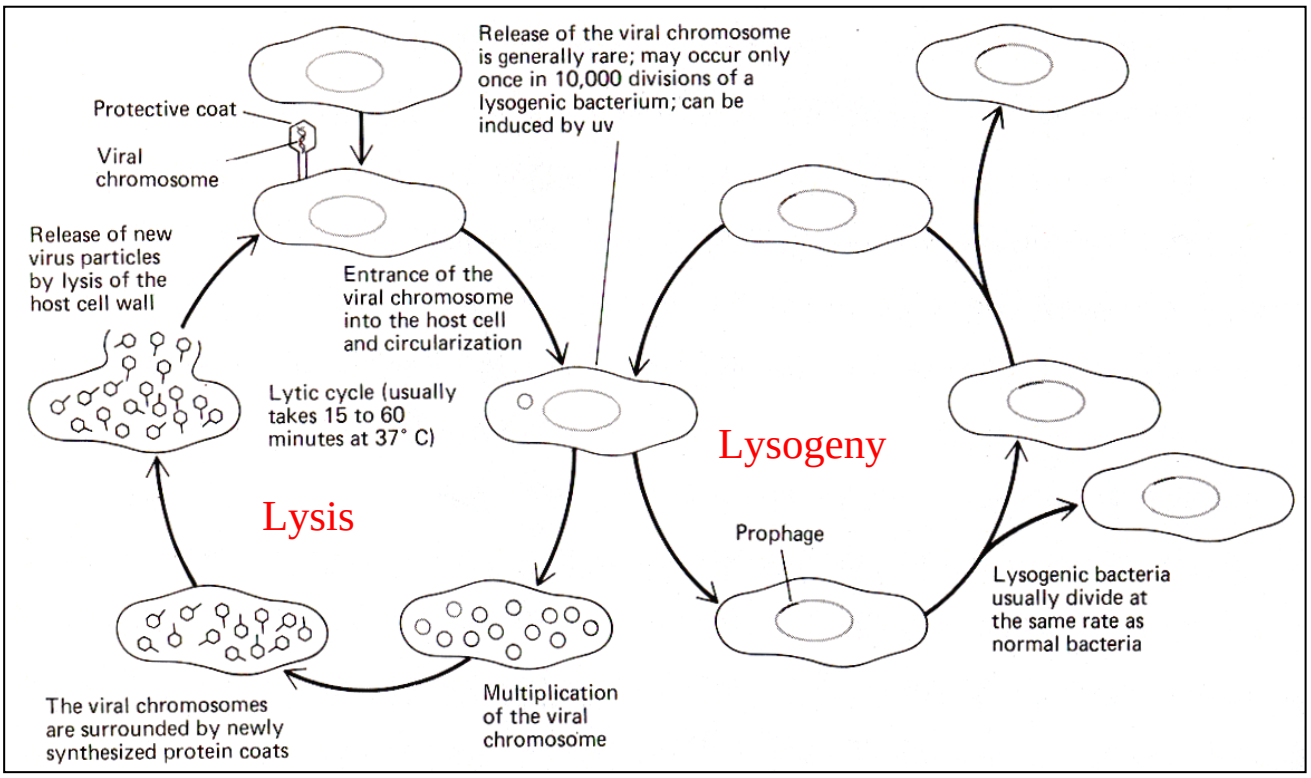


Fig 5. Phage life cycles

A phage capable of lysogeny (called a temperate phage) must necessarily address several life/death issues. First, where is the phage DNA to reside within the bacterium? Some phages replicate within the lysogenic cell as plasmids, small independently replicating units. Most, however, integrate into the bacterial genome, solving the problem of replication, since their DNA will be replicated along with the bacterium's. Lambdoid phages fall into the latter camp, integrating at a specific sites within the bacterial genome. Then there's the question of how to escape from lysogeny if a new calculation points to lysis as the better course. Lysogenic phages must be sensitive to the health of the bacterium and to environmental conditions and be ready at a moment's notice to pop out of the genome and recommence lysis. Phages that carry toxin genes are generally temperate phages, carrying the toxin gene into the bacterial chromosome when they integrate and transferring it to new hosts when they excise and undergo lytic multiplication.

III. The Scenario (revisited)

Now we can appreciate the goals of the Scenario. Bacteriophage 933W carries the gene for a potent toxin, Shiga toxin 2, which is responsible for the virulence of enterohemorrhagic *E. coli* (EHEC) strain EDL933 (the "Jack-in-the-Box strain" that caused an outbreak in the Pacific Northwest in 1993). EHEC strains carrying this phage are associated with the most severe forms of disease caused by these bacteria. What determines the bacterial targets that phage 933W can infect?

The specific receptor for phage 933W has yet to be determined. One might think that by identifying the receptor and the characteristics of the tail fibers, we could define the group of bacteria to which Shiga toxin could be easily transferred.

Unfortunately, things aren't so simple.

In addition to prophage 933W, the EHEC strain EDL933 carries at least 13 cryptic prophages (prophages that have lost some essential phage function so that they can no longer grow as phages). Seven of these cryptic prophages (CPs) contain tail fiber gene sequences with significant similarity to the tail fiber gene of phage 933W, especially at their 3' ends, but also significant differences (Fig 6).

Can these genes serve as a reservoir of variant tail fiber sequences that might expand the host range of phage 933W?

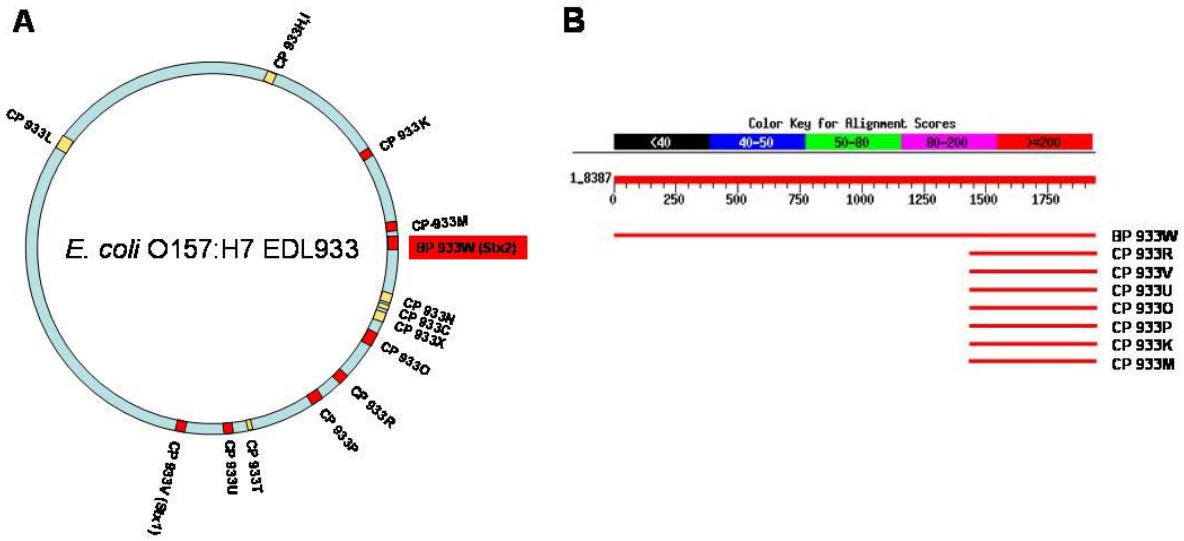


Figure 6. (A) Schematic map of the EHEC strain EDL933 genome, showing the locations of the Shiga toxin 2-encoding bacteriophage BP933W and cryptic prophages. Cryptic prophages designated by red boxes carry genes with homology to the putative tail fiber gene L0121 of BP933W. (B) Graphic output of a BLAST search of the EDL933 genome with the L0121 tail fiber gene, using default parameters, showing that a region of sequence at the 3' end of this gene is similar to sequences in 7 of the cryptic prophages.

In order to investigate this possibility, you construct a mutant strain of *E. coli* EDL933 in which the 5' end of the 933W prophage tail fiber gene, L0121, has been removed and replaced with an antibiotic resistance gene. This mutant prophage can still excise from the chromosome, replicate, package, and lyse the host, but the resulting phage particles are noninfective because they lack a functional tail fiber. You observe that there are rare phages in the lysate that ***are*** able to infect *E. coli*, however, and you hypothesize that they are the result of recombination between the now defective prophage 933W and the always defective cryptic prophage. Recombination may lead to phage DNA that has a reconstructed tail fiber gene.

How can you test this hypothesis?

In order to examine the sequence of a specific small region of the phage genome carrying the tail fiber gene, you will use the polymerase chain reaction (PCR) to amplify the DNA of interest. This process relies on the use of unique oligonucleotide primers that will anneal to specific sites on the DNA and allow the synthesis of the DNA between primer pairs (Fig 7,8; see also http://users.ugent.be/~avierstr/principles/pcr.html)

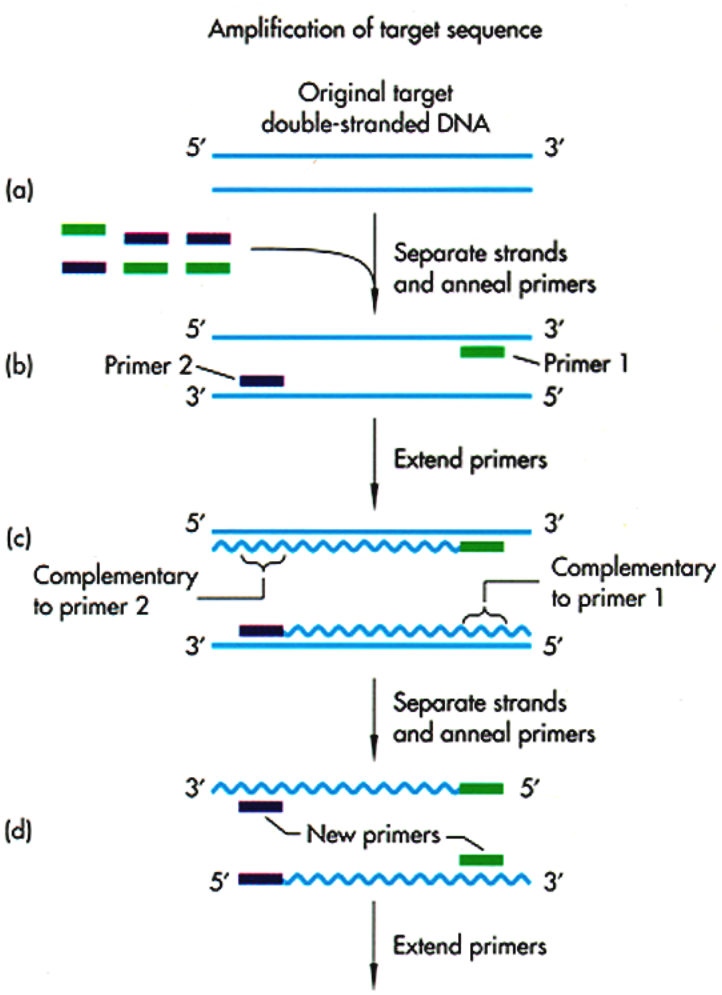


Fig 7. PCR involves synthesis of DNA from specific oligonucleotide primers complementary to the target DNA

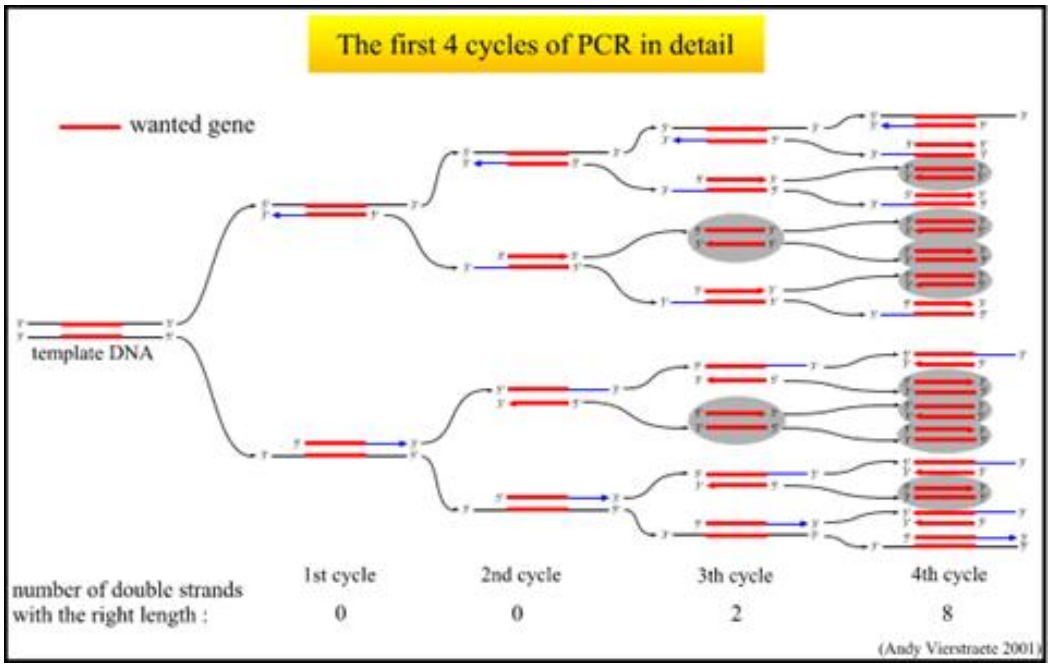


Fig 8. PCR ultimately results in the exponential amplification of those sequences that lie between the two oligonucleotide primers.

Your goal is to design oligonucleotide primers for PCR that will distinguish between the wild type 933W tail fiber gene L0121 and the CP tail fiber genes and allow you to demonstrate that the infectious phages isolated from the induced tail fiber mutant prophage are in fact recombinants.

Since recombination might take place between L0121 and any of the CP tail fiber genes, it would be convenient to find a 5' primer that matched all of the CP genes but not 933W. Likewise, it would be ideal (but not necessarily possible) to find a 3' primer that matched 933W but not any of the cryptic phage. In short, you want primers that will amplify all possible recombinant phage and only recombinant phage.