Bacteriophage genomes_Ines Staes

Slide 1

The genetic diversity of the bacteriophage population is enormous and this is reflected in the wide diversity of their genomes. In general, the nucleotide sequences of genomes derived from phages with non-overlapping host ranges rarely share sequence similarity. Because of their relatively small size and simplicity of isolation, bacteriophages were the first complete genomes to be sequenced. Today, more than 2000 unique bacteriophage genomes have been sequenced.

Slide 2

The large majority of the bacteriophages are double stranded DNA (dsDNA) viruses although the other types also exist. Phage genomes are typically linear when they are packed in the capsid. Some of the phages DNA have defined ends. This means that when you look at the linear DNA in different capsids, they will always have the same left and the same right end. For others this is not the case, for these viruses, the overall composition of the genome will be the same in every capsid, but the ends are not always the same, the DNA appears to have been linearized by opening identical circles at different sites. This is what they call circularly permuted. Additionally, the beginning and the end are often the identical sequences and this is called terminal redundancy.

Slide 3

Phage genome sizes vary enormously. The smallest are little more than 3000 nucleotides such as the ssRNA viruses of *E.coli* while the largest viruses are almost 500 kbp such as the *Bacillus Megaterium* phage G. In general, the phage DNA is usually densely packed inside the capsid at more or less similar densities; the size of the capsid thus varies depending on the genome size. Many viruses, especially DNA bacteriophages, show systematic layers of packaged DNA while others, mostly filamentous, viruses package their genome as helices surrounded by proteins. As there is not a lot of space in the capsid, the gene density of phage genomes is very high, meaning there are very few noncoding regions. The regulatory regions are often compact and occasionally the coding regions even overlap.

Slide 4

Something that is very striking when you look at phage genomes is the mosaic build-up. It seems as if the phages are constructed from different segments that have been put together like modules. A simple general explanation is that horizontal genetic exchange plays a dominant role in shaping these genome architectures. Similar groups of genes always “travel together through evolution”. Examples of this include the tail genes, or lysis genes, head genes, DNA replication genes. Bacteriophages genomes possibly also harbor the greatest genetic novelty in the biological world, in that most of their encoded genes (perhaps as much as 80%) are unrelated to known proteins, and are of unknown function.

Slide 5

Despite the fact that their genomes are orders of magnitude smaller in size compared with bacteria and other organisms, sequencing of phage genomes poses several challenges: (1) obtaining pure phage genomic material, as phages are not-self replicative, isolation of phage genomic material involves several purification steps; (2)PCR amplification can be biased as some phage genomes are notoriously rich in extreme GC content. Such extremes may pose a problem for PCR and sequencing; (3) Phage genomes are also known to contain complex genomic structures such as extremely long direct or

inverted repeats and terminal redundancies that are problematic for assembly of the whole-genome sequence from the reads and (4), in bacterial and human genomics, mapping of reads to a finished reference genome can be a powerful tool, in phage genomics this is very seldom possible due to the absence of a reference genome for any given phage. A possible solution is a blended approach of a long-read technology for scaffolding purposes combined with a large number of short reads from a second technology for efficient DNA sequencing of bacteriophage genomes.