1. Your team runs a genome sciences centre that has recently received $500,000 USD worth of sequencing funding from a Swaziland start-up to sequence an economically relevant salamander, *Necturus swazilandeus* that produces morphogens with the potential to regenerate human limbs. The Swazi researchers are interested in looking at loci conferring limb-regeneration capabilities. Using cytogenetic information, the genome size has been estimated to contain 20 billion base-pairs (Gbp). The genome is known to be highly repetitive and repeats can be as long as 6 Kbp. What sequencing platform(s) will you choose to use for sequencing this genome assuming you want 50X coverage*? How will this impact your choice of assembly paradigm [de Bruijn | OLC]? (Hint: use an equation to estimate number of sequences and cost per Mbp.)
   * The coverage must sum to 50 for any combination of technologies (e.g. 40X PacBio and 10X Sanger sequencing). At least 10X coverage is required for any one technology used.

2. You are working at the Canadian Centre for Disease Control when a viral epidemic breaks out in Vancouver. The virus is a virulent strain of beaver pox that appears to have crossed the species barrier causing infected humans to crave the taste of wood. Your team has isolated the viral particles in a blood sample and needs to sequence it immediately in order to develop an effective vaccination strategy that saves Stanley Park. The viral genome is a circular double stranded DNA molecule approaching 42 Kb in length. Your sample contains approximately one virus nucleotide per 100 thousand human nucleotides. With a sequencing budget of $80,000, what sequencing platform will you use to sequence this "metagenome" assuming you want 100X coverage? How will this impact your choice of assembly paradigm [de Bruijn | OLC]? (Hint: use an equation to estimate number of sequences and cost per Mbp.)

3. As a standard quality control step you are inspecting four reads from a sequencing run conducted at your genome sciences centre:

   ACTAGTTCT TTTGAACC TTCTATTTG TGAACCGC

Write out a de Bruijn graph representation of these reads using a k of 4. (Hint: First decompose each read into k-mer space and then find the overlaps between (k-1) mers.)