**TRACK 1 - GENOME SKIMMING.**

**ASSIGNMENT 1 – STUDY DESIGN**

In this assignment, you will need to design a project that uses Next Generation (High-Throughput) Sequencing technology to answer a biological question of your interest. In the report you should include the following items:

* 1. Your study group (i.e. organism, species, populations, clades)
  2. Give a brief background of your study group (for example: how divergent are the species/populations/individuals? how old (approx.) is your study group?)
  3. What is the overall “question” that you are trying to answer using this technology? (For example: A phylogeny of a rapid radiation of species)
  4. How many samples are you including in the final dataset? (i.e. total number of species, total number of individuals, total number of populations)
  5. What part of the chloroplast genome are you interest in? What part do you think will give you the best resolution for your “question”?
  6. Are you interested in specific genes or regions? or would you use the entire chloroplast genome?
  7. How many species/individuals are you pooling together in one sample/lane?
  8. How many lanes would you need to use in the flow cell to get sequences from all your species/individuals sampled?
  9. How many illumina runs would you need to do to get all your data?
  10. Would you get single-end or paired-end data?
  11. What is the approximate coverage that you would like to get for all your sequences?
  12. How much memory storage (hard drive space) would you need?
  13. How much RAM memory (to run analyses) would you need?
  14. How would you analyze your dataset? List the main steps you would like to do, and a specific program/script/pipeline that would be useful.
  15. Is there a “reference” available for your study group? Would you use it? Why? or why not?
  16. How long do you think it would take to finish this project? (Assume that you have already collected all your samples already, and are starting with DNA extractions)
  17. List two (2) problems that you might encounter while you are working with this project (in the lab, with bioinformatics, with your data).
  18. How would you try to solve these two specific problems?