Applied Genomics

Saul Pierotti

November 18, 2019

Course structure

- Population genetics
- Genome structure and variability in vertebrates (we may mention plants and bacteria)
- High throughput genomic platforms
- Applications of NGS
- Array comparative genome hybridization
- PLINK, genetic data analysis. How to use this software and apply some design using this tool
- Linkage analysis and genetic mapping
- QTL analysis

Examination mode

- Final exam has 2 levels
  - Preparation of a genomic project
    - A text should be written including an appropriate introduction to the problem/question that the experiment or project would like to analyse or answer, aim of the project, a section with materials and methods, expected results and impact
    - The project should be submitted to the professor one week before the interview
    - We should specify What is the aim of the project and what I'd like to solve with it
      - If it makes sense, we can undergo a discussion with him
    - The project is based on money: we'll have a budget
  - Interview based on the project submitted and other two questions
    - Only students that are positively evaluated at the first level are admitted at the second level
    - Evaluation of basic knowledge
- We get one extra point if we pass at the first attempt
- It is important to follow him
- We'll have an example of a project, the topic of the project it's up to us
- We need to choose a complex genome/organism
- Each one will have a different budget
- It's better to do the project according to what we discuss in the lectures
- It has to be something new
- The first date would be in February after Winter School and another one in March
- Near to the end of the course we'll have a test with 30 questions to test our level (it won't count for the final score)

Introduction

- Genomics is the study of genome structure and function
- The genome is the entire genetic content of an organism

- Applied genomics is the use of technologies, tools and experimental designs to analyse genome and extract information form them
- A reference genome of a species is the basis used for analyzing the genome of an individual
- We have about 2 nuclear genomes per cell, but even thousands of mithocondrial genomes
- Mithocondrial genomes can be not all equal: heteroplasmy
- The human nuclear genome is around 3 Gb, the mithocondrial genome 16.7 Kb
- Population genetics is important for this course
- Small population are susceptible to high levels of inbreeding
- Differences between population arise when there are reproductive barriers
- Effective population size is the number of individual that originated a population
  - It is a measure of inbreeding
- Sex determination can be mediate by sex chromosomes, temperature, ploidy
- Phenotype is influenced by the environment
- A phenotype is an observable charachteristic
- Comparative genomics is the study of genomic differences between species
  - It is really helpful for genome annotation
- The first draft of the human genome was completed in 2001, and the HGP was started in 1990, and the HGP was started in 1990
- 3% of human DNA is coding
- Repetitive sequences are problematic for assembling genomes
- Nuclear DNA is 99.99% identical among individuals, while mitochondrial genome is more similar
- The simplest definition of gene is "coding region"
- We can predict the phenotype of an animal just looking at the genotype (!)
- To do applied genomics I need a reference genome
- If I do not have a reference genome for my species of interest, I need to construct it or I can use one of a closely-related species
- The cost of sequencing is dropping in a way similar to Moore's law
  - Around 2008 the drop was much faster than Moore's law, thanks to NGS
- The shotgun approach does not have a particular target, it sequences everything
- Genomic data are typically stored in the cloud
- Hardy-Weinberg equilibrium
  - $\begin{cases} p^2 + q^2 + 2pq = f(AA) + f(Aa) + f(aa) = (p+q)^2 = 1 \\ p + q = 1 \end{cases}$
  - The allele frequencies refer to the current generation, while the genotype frequencies refer to the next generation
- Mendel's first law: alleles segregate with other alleles
- Mendel's second law: independent assortment
- Mendel's third law: some alleles are dominant on others
- Mendel's second law: independent assortment
- We reviewed PCR, agarose gel electrophoresys and Sanger sequencing basics
- NGS: Illumina, Ion torrent (Thermo fisher), PacBio, Nanopore
  - PacBio is going to be acquired by Illumina
  - We have short reads, therefore assembly is difficult