WEEK 3 ASSIGNMENT

Data Systems in the Life Sciences – BIOL 51000 | Fall 2 2020

Christina Morgenstern

SEQUENCE VARIATION

1. What are recombination events (examples and details of such recombination)

DNA recombination is one of the mechanisms by which genetic changes can occur that contribute to sequence variation in the individual [1]. During recombination, segments of DNA containing many bases on two molecules of DNA are exchanged with each other resulting in a new DNA molecule made up of different origins. Most notably, this happens during meiosis, the process of cell division that gives rise to the gametes (egg and sperm cells). Homologous chromosomes are lined up in pairs and swap segments of their DNA which is referred to as crossing over. This new combination of genes contributes to the genetic diversity of the offspring in the next generation [2].
Next to meiosis, recombination is also important in somatic cells where it is used to repair broken DNA strands. Such double-stranded breaks (DSBs) are sealed using a homologous DNA segment as repair template and replacing the damaged area with new material from the homologous chromosome [2].
Another example where DNA recombination is used to generate genetic diversity is within the genes of antibodies. Antibodies are proteins produced by immune cells (B cells) that can specifically detect certain antigens i.e. surface proteins from pathogens during an immune response. DNA recombination aids the generation of a variety of antibodies capable of detecting a variety of targets during a process that combines different exons [1].

2. Explanation of a single-nucleotide polymorphism

A single-nucleotide polymorphism (SNP) is a single letter variation in the DNA sequence of an individual. DNA is a biological molecule made up of a chain of four nucleotides: adenine, guanine, cytosine and thymine. If at a certain position in the DNA more than 1% of the population doesn´t carry a particular nucleotide, the position is classified as a SNP [3]. SNPs arise through the process of DNA damage which can be the result of chemicals, radiation or through the process of replication. While some of the damage can be repaired by intracellular repair machinery sometimes this safeguard mechanisms can fail and result in single base mutations [1].

SNPs can lead to diseases when the substitution happens in a gene region and changes the amino acid at this position. Sickle-cell anemia and cystic fibrosis are diseases that result from SNPs. Many times, SNPs don´t have an effect on the individual because the change in the nucleotide remains silent i.e. doesn´t change the amino acid of the protein. This is due to the degeneracy of the genetic code where amino acids are encoded by different sets of triplets providing some error-proneness. The types of SNPs that don´t affect the protein sequence are also called synonymous SNPs as opposed to nonsynonymous SNPs which lead to a change in the subsequent amino acid chain and thus impact the protein. A single SNP on its own might not have an effect on the traits or health of an individual but many SNPs together can shape the

outcomes associated with a trait or a disease. Associations between individual SNPs are assessed using genome wide association studies (GWAS) [4].

3. How to measure the relative "rate of change" of variation

The rate of change is a measure that allows us to better understand evolutionary processes. To infer evolutionary time from genetic mutations, the regions most interested are DNA segments that contain silent mutations i.e. that don´t change the resulting amino acid chain. Since the genetic code contains 64 three-base codons that encode 20 amino acids, some mutations don´t impact the amino acid sequence. The measure of DNA substitutions within a coding region that lead to an amino acid change compared to those that do not alter the amino acid sequence is an estimate for the speed of evolutionary change. When there are more silent changes than active changes this indicates that the sequence is preserved and therefor important. On the other hand, if active changes are more prevalent as compared to silent ones the rate of evolution is higher demonstrating continuous change and adaptation processes. In the human genome, genes of the immune system are subject to higher evolutionary rate which aids in the constant adaptation to new pathogens. This is referred to as positive selection whereas purifying selection stands for slow evolutionary rates [1].

4. Difference between orthologues and paralogues

When biological sequences like DNA, RNA or protein share a common ancestor they are referred to as homologous. Sequence alignments help in determining if two regions of nucleotides or amino acids share sequence similarity. A strong sequence similarity is indicative that two sequences have arisen from the same ancestral sequence. There are two variations of a homologous biological sequence: the sequence has either arisen from a duplication event or from a speciation event leading to paralogues or orthologues, respectively [5].

When a species diverges into two separate species the two copies of a certain gene in the new species are termed orthologues because they share the same common ancestor. An example is the human *PAX6* gene that has two orthologues in the fruit fly *Drosophila melanogaster* [1].

Genes that have arisen from a duplication event within a genome are termed paralogues. The Homeobox (Hox) or globin genes are an example of paralogous genes. These duplication events can shape whole genomes and shed light on genome evolution [5].

References:

[1] T. J. Stevens and W. Boucher, 'Python Programming for Biology: Bioinformatics and Beyond', *Cambridge Core*, Feb. 2015. /core/books/python-programming-for-biology/61762A9F672FDD8B2DD3FFF8773027B2 (accessed Nov. 07, 2020).

[2] 'recombination | Learn Science at Scitable'. https://www.nature.com/scitable/definition/recombination-226/ (accessed Nov. 11, 2020).

[3] 'single nucleotide polymorphism / SNP | Learn Science at Scitable'. https://www.nature.com/scitable/definition/snp-295/ (accessed Nov. 15, 2020).

[4] 'Single-nucleotide polymorphism', *Wikipedia*. Oct. 26, 2020, Accessed: Nov. 15, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Single-nucleotide_polymorphism&oldid=985613932.

[5] 'Sequence homology', *Wikipedia*. Oct. 10, 2020, Accessed: Nov. 15, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Sequence_homology&oldid=982846094.