Biology Vocabulary Notes

Siyi Wang

March 25, 2025

Introduction

This document is a collection of unfamiliar biology terms and their explanations that I have encountered during my studies. Each term is accompanied by its definition, examples, and additional notes for better understanding and revision.

Vocabulary List

1. Balancing Selection

- **Definition**: A type of natural selection that maintains genetic diversity in a population by allowing multiple alleles to coexist over long periods.
- **Example**: Heterozygote advantage in sickle cell anemia.
- **Explanation**: Balancing selection includes mechanisms like heterozygote advantage, frequency-dependent selection, environmental heterogeneity, and sexual antagonistic selection.

2. Rapid Sweep

- **Definition**: The process by which a beneficial allele rapidly increases in frequency and becomes fixed in a population.
- **Example**: The rapid spread of the lactase persistence allele in humans.
- **Explanation**: Rapid sweeps are often accompanied by selective sweeps, which reduce genetic diversity in the region surrounding the beneficial allele.

3. Polymorphic

- **Definition**: The presence of multiple forms or states of a gene or trait within a population.
- **Example**: The ABO blood group system in humans.
- **Explanation**: Polymorphism can be genetic, phenotypic, or molecular.

4. Lineages

- **Definition**: A series of organisms, populations, or species descended from a common ancestor over evolutionary or genetic time.
- **Example**: The lineage relationship between humans and chimpanzees.
- **Explanation**: Lineages can refer to species lineages, gene lineages, cell lineages, or cultural lineages.

5. Synonymous

- **Definition**: A synonymous mutation is a base substitution in a DNA sequence that does not change the encoded amino acid. This occurs because of multiple codons can encode the same amino acid.
- **Example**: Codons "GAA" and "GAG" both encode glutamic acid (Glutamic Acid). If a mutation changes "GAA" to "GAG", this does not alter the protein's amino acid sequence, and is therefore a synonymous mutation.
- **Nonsynonymous Mutation**: The codon "GAA" encodes glutamic acid (Glutamic Acid), but if mutated to "GTA", it now encodes valine (Valine). This mutation changes the protein's amino acid sequence, and is therefore a nonsynonymous mutation.

6. Comparison Table

Feature	Positive Selection	Purifying Selection
Definition	Increases the frequency of beneficial mutations.	Removes or reduces the frequency of harmful mutations.
Target	Beneficial mutations.	Harmful mutations.
Outcome	Drives adaptive evolution.	Maintains functional stability of genes.
dN/dS Ratio	dN/dS >1.	dN/dS <1.
Examples	Lactase persistence gene, malaria resistance genes.	Highly conserved genes.

Table 1: Comparison between Positive Selection and Purifying Selection

Ideally, this involves aligning homologous gene sequences from two or more species, ensuring codons are precisely matched.
1. Calculating the number of synonymous and non-synonymous sites within each codon
2. Counting the synonymous and non-synonymous mutations
3. Using a specific model to calculate dN and dS ???
4. Determining selective pressure by computing the dN/dS ratio

Statistical significance is assessed using a Z-test of

$$\frac{dN/dS - 1}{\sqrt{\text{Var}(dN/dS)}}$$

If $|z|$ >1.96 (at the 0.05 significance level), the null hypothesis of neutral evolution (dN/dS = 1) is rejected. R code: dnds(x, code = 1, codonstart = 1, quiet = FALSE, details = FALSE, return.categories = FALSE)

7. MHC (Major Histocompatibility Complex)

- **Definition**:A group of genes encoding proteins responsible for presenting antigen fragments on the cell surface for recognition by immune cells (such as T cells).
- **Function**: Antigen Presentation–Displaying antigen fragments to T cells.

8. Polymorphic

- **Definition**: The existence of multiple different forms or states of a gene or trait within a population. These different forms can be variations in gene alleles, DNA sequence variations, or differences in phenotypic characteristics (such as color, morphology, etc.).

9. Basic Terms

- **Haplotype**: A haplotype is a group of closely linked genes or genetic markers located on the same chromosome that are typically inherited together. A haplotype can be viewed as a combination of alleles within a chromosomal region.
- **Base Pair**: A base pair is the fundamental unit of the DNA double helix structure, formed by two complementary nucleotides (A-T or C-G) connected by hydrogen bonds.
- **Locus**: A locus is a specific location in the genome, which can be the position of a single gene, a SNP, or other genetic markers.

10. Neutral Variant

- **Definition**: A genetic variation that does not have a significant impact on an organism's fitness. These variants are typically not influenced by natural selection, and their frequency changes in a population are primarily determined by genetic drift.

11. Genotype

- **Definition**: The genetic composition of an organism, typically represented by the alleles of genes. It determines the genetic characteristics of an organism and influences the phenotype through gene expression.
- **Example**: Human blood type genes (ABO blood type system AA or AO $\rightarrow$ Type A blood; BB or BO $\rightarrow$ Type B blood; AB $\rightarrow$ Type AB blood; OO $\rightarrow$ Type O blood; Here, the genotype (such as AO, BB) determines the phenotype (such as Type A blood, Type B blood). Explanation: Homozygous: identical alleles, such as AA, aa. Heterozygous: different alleles, such as Aa.

12. Admixture

- **Definition**: when two or more populations with different genetic backgrounds experience gene flow, resulting in descendants who carry genetic material from multiple ancestral populations.
- **Example**:The genetic mixing between Europeans and indigenous Americans led to the diverse genetic composition of today's Latin American populations.

13. Confounder

- **Definition**: An additional variable that simultaneously affects both the independent and dependent variables, potentially making the relationship between them appear false or overestimated/underestimated.
- **Example**:Coffee consumption (X) may be associated with heart disease (Y). However, if smoking (Z) is ignored: smokers typically drink more coffee (X). Moreover, smoking itself increases the risk of heart disease (Y). Here, smoking is a confounder that affects our understanding of the relationship between coffee consumption and heart disease.

14. Race and Ancestry

Feature	Race	Ancestry
Nature	Social construct	Biological concept
Based on	Physical traits & cultural identity	Genetic markers
Variability	varies by society	determined by DNA
Use in Studies	Sociological analysis	Population structure
Accuracy in Biostatistics	Less precise, prone to confounding	More precise for genetic analysis
Example	"Hispanic," "Asian," "Black"	"50% East Asian"

Table 2: Comparison of Race and Ancestry in Biostatistics

15. Admixture

- **Definition**:The genetic exchange between populations with different genetic backgrounds, resulting in individual genomes containing DNA fragments from multiple ancestral sources.
- **Example**:This phenomenon can occur in human, animal, and plant populations, typically driven by factors such as migration, population merging, colonization, and wars.

16. PCR Amplification

- **Definition:** PCR (Polymerase Chain Reaction) amplification is a molecular biology technique used to make millions to billions of copies of a specific DNA sequence, starting from a very small amount of template DNA.
- **Example of Use:** In paleoanthropology, PCR amplification is used to amplify small fragments of ancient DNA (aDNA) extracted from fossil remains, such as Neanderthal bones. This allows researchers to sequence and analyze the genetic material, even when the original DNA is highly degraded and scarce.

17. Adaptive Introgression

- **Definition:** Adaptive introgression refers to the transfer of beneficial genetic material from one species or population into the gene pool of another through hybridization and subsequent backcrossing, followed by natural selection retaining the advantageous alleles.
- **Example:** One well-known example is the introgression of Neanderthal DNA into modern humans. Certain Neanderthal alleles related to immune response (e.g., against pathogens) and skin pigmentation have been found in non-African human populations and are believed to have provided adaptive advantages in non-African environments.

Conclusion

I will continue to add more terms and apply this knowledge to practical problems in the future.