Center for Computational and Theoretical Biology

Master Thesis

Selection on loss-of-function variants

Master Biosciences
Julius-Maximilians-Universität Würzburg

submitted by Laura Steinmann
Supervisor PD Arthur Korte

March 11, 2022

Supervisor PD Arthur Korte

Second reviewer Prof. Dr. Dirk Becker

Date of Submission, office stamp

Zusammenfassung

Abstract

Contents

1. Introduction

To survive in an ecosystem organisms need to adapt to the specific abiotic and biotic factors around them. This is particularly important for plants since they can not change their habitat during their lifespan. Adaptation can operate on different levels of the organisms from the adaptation of protein function to modifications in cell functions or whole tissue functions. But the underlying fundamental process that drives adaptation is the modification of the genetic material DNA.
Modifications on DNA level are called mutations and can be categorized in four different groups. A point mutation, which is the first group of mutations, is concerning just one single base pair. These changes the DNA sequence on one single point and therefore lead to changes of the transcribed mRNA. The second class of mutations are insertions. These insert a new sequence into the former DNA sequence and thus elongate it. These insertions can have a variety of length from one single bp to a few hundred bp and even longer sequences can be added. The contrasting class of insertions are deletions, which lead to a reduction of base pairs in a variety of lengths in the former DNA strand. The last category of mutations are duplications. As implied by the name regions of the DNA get duplicated and inserted at a different position. The regions can be copied abnormally one or even more times.
These mutations are first of all just changes in the DNA but indirectly they impact all the resulting processes. The DNA first gets transcribed into mRNA. This process is not affected by the evolved mutations. But after the transcription the mRNA gets translated into proteins. In this step the impact of the mutations appear since a triplet codon gets translated into a specific amino acid, which can change due to the mutation in the DNA. The mutations can have different effects on the codons. The insertions or deletions lead mostly to a shift in the reading frame. But also point mutations can have different effects on the translated codon and therefore on the functionality of the resulting protein. As it was known that 23 amino acids exists and that the knowledge arises that a codon includes 3 bp the codons could get deciphered from Nirenberg et al. [1]. The gain of knowledge deciphering this genetic code showed that not one codon is directly linked to an amino acid rather that there are multiple codons that can initiate the same amino acid, which is showed in Figure 1.1 a). For the topic of point mutations we have three possible outcomes resulting after such mutation occur. The first outcome is a silent mutation, which is the less severe one and does not change the protein at all. Here a base pair changes but the translated codons provoke the same amino acid. An schematic representation of synonymous mutations can be found in Figure 1.1 b). As shown in Figure 1.1 c) a mutation could also be a non-synonymous mutation where the modified codon leads to an exchange of the amino acid type. The prediction of how severe these non-synonymous mutations are is hard to do without further analysis. It depends on many circumstances for example if the location is a catalytic region or what the exchange amino acid is. Such a mutation can be very unnoticeable or on the other hand making the protein non-functional. The last kind of point-mutations are premature stop codons. They represent a mutation of a usual codon to a stop codon. This means most of the time drastic changes in the protein functionality since the resulting protein is truncated. Thereby premature stop codons represent the most interesting kind of mutations for understanding adaptation in plants since they can be recognized in the DNA level and they should lead to an truncated and non-functional protein.

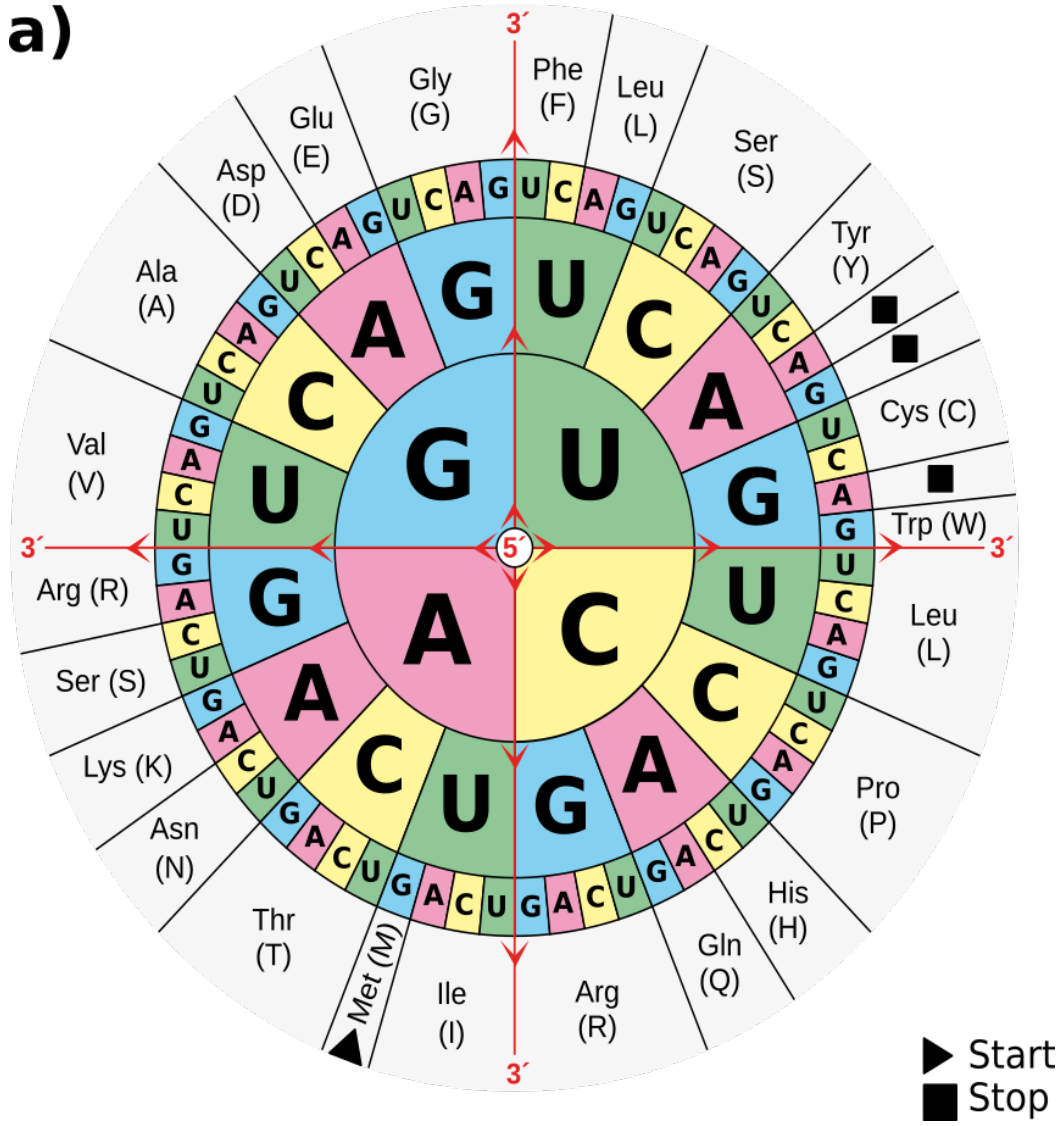

Fig. 1.1.: Fundamental concepts of Translation of DNA into the Protein
a) The genetic code (image taken from Bresch and Hausmann [2])

2. Material

List of Figures

B. Bibliography

[1] M Nirenberg et al. "RNA codewords and protein synthesis, VII. On the general nature of the RNA code." In: *Proceedings of the National Academy of Sciences of the United States of America* 53.5 (May 1965), pp. 1161–1168. ISSN: 0027-8424. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC301388/` (visited on 03/11/2022).

[2] Carsten Bresch and Rudolf Hausmann. *Klassische und molekulare Genetik*. Springer-Verlag, 2013.

Affirmation

I hereby confirm that my master thesis entitled Selection on loss-of-function variants
is the result of my own work./
I did not receive any support from commercial consultants. /
I have given due reference to all sources and materials used in the thesis and have listed and specified them. /
I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form. /
I agree that the thesis can be checked for plagiarism also by using a software./

Hiermit erkläre ich an Eides statt, dass ich die Masterarbeit mit dem Titel Selektion der Loss-of-Function Varianten
eigenständig und eigenhändig angefertigt habe.
Ich habe keine Unterstützung kommerzieller Berater erhalten.
Ich habe alle in der Arbeit verwendeten Quellen und Materialien ordnungsgemäß zitiert, aufgelistet und spezifiziert
Ich erkläre, dass die vorliegende Arbeit weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.
Ich bestätige, dass die Thesis auch mit Hilfe einer Software auf Plagiat untersucht werden kann

Würzburg, March 11, 2022 Laura Steinmann