On the trail of viruses: understanding virus research

Solution worksheet B: Sequence analysis

Mutation analysis of the SARS-CoV-2 virus using MEGA

Task 1

a) Take a closer look at the sequences arranged one below the other. Describe which codon they each begin with and explain the significance of this codon for translation.

They start with ATG. This codon is both the start codon (translation begins here) and also codes for the amino acid methionine.

b) Now scroll slowly to the right and take a closer look at the alignment. Describe in bullet points what you notice and explain why this might be the case.

From position 200 onwards, the sequences are increasingly poorly aligned due to various mutations occurring.

Task 2

a) Now scroll slowly through the sequences again and look at the alignment. Explain what has changed.

Mutations were detected and the sequences aligned accordingly. The alignment is greatly affected by insertions/deletions, especially if the corresponding gaps are not taken into account.

b) Give an example of each of the three different types of point mutation (substitution, deletion and insertion) that can be found in these sequences. Use the format C56G (cysteine at position 56 is changed to glycine).

There are two possible solutions here. If the pupils evaluate the nucleotide sequences, the possible results can be found in table 1. If the pupils evaluate the protein sequences, the possible results can be found in table 2.

Table 1: Possible results based on nucleotide sequences

Variant	Mutation		
	Substitution	Deletion	Insertion
Delta	C56G, C57A, A467G, G468C, T1364G, C1442A, C1443G, C1455T, C2050A, C2052G, C2052A, G2857A	Δ469–474	-
Omikron	C200T, C201G, C284T, G425A, G1025A, T1120C, C1121T, T1126C, C1128T, C1133T, G1260C, C1329G, G1345A, T1438A, C1439A, C1442A, C1443A, A1460C, G1461C, C1649G, C1650G, C1972T, C2046G, C2051A, C2301G, G2395T, C2577T, G2871C, C2916G, C2950T	Δ205–210, Δ427–435, Δ631–633	Ins642 GAGCCCGAG

Table 2: Possible results based on protein sequences

Variant	Mutation		
	Substitution	Deletion	Insertion
Delta	T19R, R158G, L452R, T478K, P681R, D950N	Δ156–157	–
Omikron	A67V, T95I, G142D, L212I, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	Δ69–70, Δ143–145, Δ211	ins214EPE

Task 3

Examine the sequence alignment sites listed under a) to d) in more detail and name which variant is mutated and how. Describe the variant, the type of point mutation, the corresponding nucleotides and the resulting amino acids. Use the code column at the end of the worksheet for the latter.

TEACHER INFO: If pupils have problems translating correctly due to an incorrect codon, you can tell them that if the sum, when divided by three, is ***. $\overline{3}$ or ***. $\overline{6}$, they are at the first or second position of the codon.

a) 205–210

Omicron, deletion of CACGTC or histidine H & valine V

b) 642/643

All 640–642 CGC → arginine (R); Omicron at position 643, insertion of GAGCCCGAG or glutamic acid/proline/glutamic acid (EPE)

c) 1501–1503

Omicron, substitution, AAC→ TAC, asparagine N → tyrosine Y

d) 2041–2043

Delta, substitution, CCC → AGA, proline P → arginin R and Omicron, substitution, CCC → CAC, proline P → histidine H

Task 4

a) In frequently used medical PCR tests, several sites in the genetic material of the virus are compared. The mutation Δ69-70 (in the protein sequence) falls within one of the binding sites for PCR primers. Explain how the PCR results of someone infected with the omicron variant differ from someone else infected with the delta variant.

At this point, the primer will only bind to the Delta variant, but not to the Omicron variant. The band pattern in gel electrophoresis would therefore only show a band in the person infected with the Delta variant.

Addition 1: It is pure coincidence that the Omicron variant does not produce a PCR product with this primer. The test was not designed to distinguish between variants. For diagnostic purposes, it is beneficial that the test examines three or more genes. Otherwise, the Omicron patient would have tested negative.

Addition 2: This mutation is also important for the immune response, as existing antibodies (after vaccination or previous infection) appear to bind to this site. The loss of these two amino acids means that the existing antibodies can no longer bind as effectively. This results in a weakened immune response.

Source: https://insidecorona.net/the-new-mutation-of-sars-cov-2/

b) Analyse the side chains of the two amino acids in the mutation N501Y. Describe the chemical properties and differences.

N = asparagine (polar/neutral), Y = tyrosine (polar/neutral). The chemical properties of the two amino acids are identical. Their side chains are both neutrally charged and polar. However, the side chain of tyrosine is larger because, unlike asparagine, it contains a cyclohex-2,4,6-trien molecule. This leads to possible steric changes.

Addition: The N501Y mutation is located at the contact surface between the ACE2 receptor of the human cell and the spike protein. This mutation enhances the binding of the spike protein to ACE2. This enables the virus to penetrate the cell, resulting in higher transmissibility.

Source: https://insidecorona.net/the-new-mutation-of-sars-cov-2/

c) Examine the side chains of the amino acids of P681R at Delta and of P681H at Omikron. This site is close to a cleavage site that is important when the virus enters the host cell. Describe the chemical properties and differences. Discuss whether, and if so, what effects are to be expected for the infection

P = proline (hydrophobic), R = arginine (basic), H = histidine (basic). On the one hand, a neutral, non-polar amino acid is changed into a basic one (applies to arginine and histidine). This can have consequences for intermolecular interactions. Non-polar amino acids can only interact with other non-polar amino acids via Van der Waals forces. Basic amino acids, on the other hand, can form real chemical bonds, in this case ionic bonds. Secondly, the size of the side chain also changes here, leading to steric changes.

Addition: A research group at the University of Texas was able to prove that the P681H mutation in the Omicron variant enabled the virus to infect cells more quickly and efficiently, leading to the rapid spread of this variant.

Sources:

- https://www.spektrum.de/news/sars-cov-2-die-mutation-die-delta-so-ansteckend-macht/1915573
- https://www.nature.com/articles/d41586-021-02275-2

Additional task

Compared to viruses, higher animals have a much larger genome. Some genes are present more than once (through duplication or diploidy).

a) Hypothesise what the multiple presence of genes may mean for the survival chances of an individual with a mutation.

If a gene is altered by mutation, its function may be lost or impaired. If you have multiple copies of a gene, you have a spare copy that still contains the original information. This allows evolution to develop two genes with different functions from one gene.

b) Explain whether a difference in the frequency of mutations (surviving to the next generation) is to be expected between viruses and higher animals such as humans.

Yes, but it's complicated. Mutation is much more important/dangerous for a virus because there is only one copy of the gene, and each gene has an important function to perform (less overlap than in higher organisms). But viruses also have short generations, so successful mutations can establish themselves quickly. Mutations are therefore riskier for viruses than for humans, but at the same time they spread much faster.

Homework

Task 1

a) Find out about the topic of artificially mutated viruses and draw up a list of pros and cons for a discussion group. Make a note of the sources you use.

Note: It would probably be important to reiterate the appeal to only use trustworthy sources.

b) Find out about the advantages and disadvantages of experimenting with artificially modified and thus mutated viruses. Choose a side of the argument and prepare to present and defend it in a discussion round. Make a note of the sources you use.

We would like to encourage discussion on this topic and have therefore examined a number of possible aspects. The list of links is not exhaustive and was last checked for accuracy on 10 November 2023.

Aspect	**Links**
Biological warfare	Life science: AI engineered viruses Global bio defense: Ebola as bioweapon Uni Freiburg: Insect born Viruses Lieber Institute: China violate biological warfare convention Techno-Science.net: Deadly viruses missing from lab
Medical science	Helmholtz: Bacterial weapons against viruses Newswise.com: Viruses that heal Yale news: Hunting for 'good' viruses BBC: I found a bacteria-eating virus
Ethics	Harvard Medical School: Bioethics and The Future: Can Progress Be Tamed? Wold Medical Association: WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Participants Frontiers: Ethical challenges in research involving children affected by armed conflict National Institute of Health: Guiding Principles for Ethical Research

Task 2

Find out more on the internet about PCR tests and their use to detect a corona infection.

a) Briefly summarise the procedure of a PCR test.

A PCR test can be used to reliably detect the presence of pathogens from a very small sample, e.g., a swab of the nasal mucosa. Polymerase chain reaction is a standard laboratory method used to amplify DNA. In the case of the SARS-CoV-2 virus, this involves genetic material from the virus, which is found in the nasal swab. The amount is too small for actual detection, so PCR is used to increase the amount. By using different primers during PCR, it is possible not only to detect a coronavirus infection, but also to determine which virus variant the person being tested is infected with. PCR is a very sensitive method, as only very small amounts of DNA are required. In addition to detecting infections, PCR is also used in forensics and food testing.

The steps for detection are as follows:

- Sample collection
- Sample preparation
- Polymerase chain reaction (PCR)
- Gel electrophoresis
- Evaluation of gel electrophoresis

b) Explain why several genes are taken into account in the PCR test.

The starting points for PCR are the primers. They are between 18 and 30 nucleotides long and bind specifically to a certain position on the DNA. If there are mutations in the binding site for the primers, they may bind less well or not at all. In our example, the following scenario would then be possible: If the PCR is only performed with specific primers for the Omicron variant, but the sample being tested contains the Delta variant, the PCR test result will be negative, even though SARS-CoV-2 viruses are present. For this reason, several primers for different variants are always selected for PCR so that an infection can actually be detected regardless of which variant led to the infection.

Acknowledgements

This resource has been sponsored by the Joachim Herz Stiftung and produced by the European XFEL.