A newly emerged human virus is hitting the world. Everywhere around the globe, governments demand that people keep physical distance from each other, entire countries are in lockdown, borders are closed and flight connections halted. Researchers are working at full speed to find cures and vaccines. What is this new virus that has changed our lives so drastically in such a short time? How did it arise? How can it spread so easily? There is a lot that we do not yet understand, but here is a short summary of what we think we know and how we can fight back.

When we speak about this new virus, we often just say ‘the coronavirus’. However, there is not just ‘a’ coronavirus. The word ‘coronavirus’ refers to a family of related viruses, and there are thousands of them. They are large, enveloped RNA viruses with diameters of 70 to 120nm. (‘Enveloped’ means that their capsids are wrapped in a lipid layer.) They have the largest genomes of all RNA viruses (26 to 32kb). Coronaviruses owe their family name – which is the Latin word for ‘crown-viruses’ – to their appearance: spikes on their surface make them look as if they are surrounded by a crown.

Coronaviruses infect a wide range of mammals and birds such as cats, dogs, chickens, and even beluga whales. Bats are a particularly common host of coronaviruses (without suffering symptoms), and as you will see below, they played an important role in the current outbreak.

The proper name of the coronavirus causing the current pandemic is SARS-CoV-2, which stands for ‘severe acute respiratory syndrome coronavirus 2’. It is the seventh (known) coronavirus that infects humans (see Box on human coronaviruses on the next page). All of them cause mainly infections of the respiratory tract or the digestive system.

The disease caused by the new coronavirus is called Covid-19, which stands for ‘coronavirus disease 2019’. In many people, the virus causes mild symptoms or none at all. But for others – especially for the elderly and for those whose health is already compromised for other reasons – it can be life-threatening. In severe cases, patients are unable to breathe and require ventilation. Even though people in risk groups are more likely to be strongly affected by the disease, anyone – even healthy and fit young individuals – can develop extremely severe symptoms and even die.
The seven coronaviruses known to infect humans are called HCoV-OC43, HCoV-229E, HCoV-NL63, HCoV-HKu1, SARS-CoV-1, MERS-CoV, and SARS-CoV-2 (the new one). The first four have been circulating in the human population for a long time. They usually cause mild symptoms and are responsible for 15%-30% of cases of the common cold each year. SARS-CoV-1 emerged in the Guangdong province of China in 2002 and sparked a violent pandemic outbreak. The mortality rate was around 10% on average. For people older than 65 years, however, it was even as high as 50%. Overall, slightly more than 8000 people were infected and almost 800 of them died. The last cases were detected in 2004. MERS-CoV (Middle East Respiratory Syndrome coronavirus) first appeared in 2012 on the Arab Peninsula. Like the two SARS strains, it causes severe symptoms, though it has an even higher mortality rate of 35%. Fortunately, unlike the other human coronaviruses, it does not transmit well between humans. The new human coronavirus SARS-CoV-2 emerged in late 2019. Analysis of its RNA sequence revealed that it is closely related to SARS-CoV-1 (hence its name). Disease symptoms range from very mild to very severe, and many infected individuals remain entirely asymptomatic. The mortality rate is lower than that of SARS-CoV-1 or MERS – probably around 1-2% for symptomatic infections (but this number is still a vague estimate due to the many undetected cases). Unlike MERS, the virus is easily transmitted between humans, leading to a large number of cases.

In the following, you will read how researchers think this new virus emerged, how it infects us, how we can test whether someone is infected by the new virus, what antibody tests are and why they are important.

The emergence of SARS-CoV-2

Most, if not all, human viruses derive from animal viruses (see article “A glimpse into the world of human viruses”). This is also true for the novel coronavirus SARS-CoV-2, which jumped from bats to humans possibly via an intermediate host. Let us look at this a bit more closely.

Crossing species boundaries

Viruses are normally adapted to their specific hosts, which means they only infect individuals of a particular species. A given virus cannot simply infect hosts of a different species. The reason is that the proteins on cell surfaces and within cells differ between species. The viral tools that perfectly match the proteins of one host species do not work – or at least not well – on the proteins of another. This means that the virus cannot enter the host cell or is unable to hijack its machinery to become replicated. However, viruses can evolve and adapt to infect hosts of another species.

![Fig. 2: Schematic of a coronavirus. The red proteins are the spike (S) glycoproteins that the virus uses for cell attachment and entry. Under the electron microscope, they appear as a 'crown' around the virus (see Fig. 1). Credit: CDC/Alissa Eckert, MS; Dan Higgins, MAMS](https://www.cdc.gov/vhf/merck/sars코로나바이러스.html)
room for mutations to occur. Yet, their mutation probability is actually lower than that of other RNA viruses since they possess repair mechanisms, which most RNA viruses do not. However, when two different coronaviruses infect the same cell, there is a high probability that they recombine with each other – i.e. swap part of their DNA. This creates a new virus that may have quite different features. Their overall abundance plus ample time to mutate and recombine results in great genetic diversity in coronaviruses, making it likely that one of them matches a new host at least well enough to somehow proliferate.

Another – and very important – reason has to do with the viral protein that the coronavirus uses to attach to the host cells and to enter them. Host cell attachment and entry are mediated by the so-called spike (S) glycoprotein that is part of the viral envelope. (This spike protein gives the virus its crown-like appearance.) The coronavirus spike (S) glycoprotein is very tolerant to changes, which means that the corresponding gene can acquire many mutations without compromising the viability of the virus. The existing variation in this glycoprotein and its variability means that there likely exists a variant that can attach to a new host and that subsequent further adaptation to the new host is possible. For example, we know from the first SARS epidemic in 2002-2004 (see Box on the seven human coronaviruses) that the glycoprotein sequence was under positive selection while the virus adapted to the human host.

There is yet another factor facilitating coronavirus host jumps. To understand this, we not only have to look at the viral spike (S) glycoprotein but also at the host cell receptors (that is, proteins at the host cell surfaces) to which it binds to enter the cell. Even though distinct strains of coronaviruses may use different receptors for the same host, the two SARS coronaviruses (and related coronaviruses that infect animals) all use the receptor known as ACE-2. Laboratory experiments have demonstrated that SARS-CoV-2 is able to bind to the ACE-2 receptors of other species in addition to humans. We have also seen cases of cats and tigers infected with SARS-CoV-2 outside of the lab. This means that this receptor is quite similar across many species, which in turn means that jumping to another host is easy for coronaviruses that use this receptor for cell entry. It wouldn’t even need to change very much.

The origin of SARS-CoV-2

Bats harbor a huge variety of coronaviruses and have been identified as a reservoir from which jumps into other species occur. As far as we know, all three coronaviruses that caused epidemics in the 21st century – SARS-CoV-1, MERS-CoV, and SARS-CoV-2 – have their origins in bats. For SARS-CoV-1, palm civets likely acted as intermediate hosts, where selection was still ongoing at the time it jumped to humans. For MERS-CoV, the virus jumped first from bats to dromedary camels, and those form the reservoir from which the virus jumps to humans. Camels themselves may develop symptoms of a respiratory tract infection, such as fever or a runny nose, but may also remain asymptomatic. Spillover events of MERS-CoV from camels to humans still occur, resulting in severe disease with high mortality rates.

How did SARS-CoV-2 jump over to humans? We don’t really know yet. It may have jumped directly from bats to humans, or it may have been transmitted from bats to an intermediate host and from there to humans. The way researchers reconstruct the likely pathway is by comparing the RNA sequences of human and animal viruses. From this, we know that SARS-CoV-2 is closely related to a bat coronavirus and assume that it is not a recent recombinant (i.e. combination) of different coronaviruses, but no conclusive picture about the order of events and possible intermediate hosts has emerged so far.

Whatever the pathway, we know that the epidemic started in late November or early December in Wuhan in China.

Spread of the virus

On December 31, 2019, Chinese authorities informed the WHO of cases of pneumonia, for
which they did not know the causal pathogen. The pathogen was identified as a coronavirus on January 7, and on January 10, its genome was published.

Starting with the local outbreak in Wuhan, the virus quickly spread around the world. On March 11, the WHO declared COVID-19 a pandemic. By that time, 114 countries were affected, almost 120,000 cases had been identified, and more than 4,000 people had died. Six weeks later, the number of detected cases has reached more than 2.5 million and more than 190,000 people have died of COVID-19.

How does the virus spread so quickly?

Further evolution of SARS-CoV-2

Of course, the virus continues to mutate as it spreads through the human population. But although coronaviruses are generally quite evolvable, researchers found that at the timescales of the pandemic, SARS-CoV-2 evolves rather slowly. It seems to accumulate mutations at a slower rate than the human influenza virus. Also, at somewhat longer timescales, it is not expected to evolve nearly as rapidly, since unlike the influenza virus, it does not have a segmented genome that it can readily reshuffle with other coronaviruses (see the article “A glimpse into the world of human viruses” to read about the rapid evolution of the influenza virus). This is important and advantageous since it means that the development of effective vaccines is likely possible. A vaccine might not provide life-long immunity as it does for measles or rubella (not only because the virus may change but also because immunity may wane). But it will hopefully provide very good protection for an extended period of time, and the development of updated vaccines can keep up with the pace of evolution of the virus.

Without any countermeasures in place, each infected person infects on average 2-3 other people. This leads to the explosion of cases we have observed in Italy, Spain, France, Germany, etc.: the first patient infects 3 other people; these infect 9 others who in turn infect 27 new people. In the next step, the number of new infections is already 81. You can see that the increase in cases speeds up! Numbers that grow in this way are said to grow exponentially. No healthcare system can keep up with this.

The virus is transmitted from human to human mainly via droplets we release into the air while breathing, speaking and especially when we cough or sneeze. This is why physical distancing is so important to slowing the spread of the virus through the human population. The size of these droplets varies, which results in different outcomes. Large ones fall to the ground quickly. On the other hand, the water in small droplets evaporates before they reach the ground, leaving behind minerals, proteins, and viral particles. These tiny water droplets and their remains are called aerosols. These can remain in the air for a long time, and there is a lot of debate how much they contribute to the transmission of SARS-CoV-2. We can potentially also catch the virus when we touch contaminated surfaces and afterwards touch our noses, mouths, or eyes, transporting the virus to tissues where it can cause an infection (the virus cannot infect the skin cells of our hands). Therefore, handwashing is one of the measures to protect ourselves: the soap destroys the lipid envelope of the virus, making it unable to do any harm.

You may wonder why you have to adhere to physical distancing even though you and your friends are perfectly healthy. In fact, people are probably infectious for some time before having symptoms. Moreover, many people do not develop any symptoms at all – but they are still infectious and can transmit the virus to others. This is the reason this coronavirus pandemic is so difficult to contain. With SARS-CoV-1, people only became contagious after they showed symptoms. Recognizing and isolating people with symptoms is relatively easy. This is one of the reasons
SARS-CoV-1 could be effectively contained. Recognizing and isolating people who are infected but perfectly healthy, in contrast, is hard. Therefore, we all have to keep a safe distance from others right now.

The two SARS coronaviruses spread differently because of differences in the cells they infect. SARS-CoV-1 infects lung cells, causing severe illness. SARS-CoV-2, in contrast, proceeds in two steps, making it both very transmissible and very dangerous. It first infects cells of the throat, causing no or mild symptoms. From the throat, it can be transmitted easily to other hosts – similar to the four common coronaviruses, which also infect the cells of the throat and cause common colds. Later, in a fraction of infected individuals, the viral infection expands further down into the lung, making people seriously ill.

Mathematical modeling of pandemics

Mathematical modeling plays an important role in understanding how the virus spreads and how we can best stop it. First, mathematical analyses are needed to interpret epidemiological and sequencing data. For example, by comparing SARS-CoV-2 sequences from infected patients and by calculating the probabilities of different scenarios of transmission, mutation, and spread, researchers found out that the virus was spreading undetected near Seattle in the US. In addition, the insight that every infected infects on average 2-3 other people was obtained very early on using mathematical modeling. Likewise, researchers try to estimate the number of undetected cases through models. Moreover, mathematical models are used to explore the likely effects of control measures and their relaxations.

Physical distancing is, of course, not a permanent solution. But it helps in many ways. First, by lowering the number of new infections, it lowers the number of patients who get seriously ill and need treatment in an intensive care unit in a hospital. This is of utmost importance since these spaces are limited. Second, we gain the time we need for the development of cures and vaccines, faster testing technologies, and antibody tests (see below). Last, if we manage to slow down the spread so much that each person infects less than one other person on average, we can drive the virus to extinction. Globally, this will not be possible but locally, it can be achieved.

How does testing work?

It is very important to know whether a person is actually infected with SARS-CoV-2 or not. If we know that a person is infected, the person can be quarantined until they have recovered and are no longer infectious. Through contact tracing, people who met this person and may have contracted the virus can be identified and quarantined as a precaution as well. However, many symptoms (e.g. fever or coughing) are not specific to COVID-19, especially in mild cases. Moreover, as we said above, many people do not develop symptoms at all. Diagnosis based on symptoms is thus not always possible, which means we need reliable laboratory diagnosis.

Detecting the virus

The tests developed to detect the virus are based on a technique called real-time reverse-transcriptase polymerase chain reaction (real-time RT-PCR). The goal is to make the viral RNA visible if it is present in swabs taken in people’s mouths or noses. For this, the viral RNA – if present – is first reverse-transcribed into (single-stranded) DNA. Then a segment of this DNA is replicated over and over again to obtain millions of copies (see Box on RT-PCR for details). Through an ingenious construct, every replication event leads to the release of a dye molecule that glows if irradiated with light (see Box on real-time RT-PCR). Therefore, the increase in the number of DNA segments is accompanied by an increase in the glow intensity. However, this only happens if the specific viral RNA (and hence DNA) was present in the swab. In this way, we know whether the patient is infected or not.
Let us assume we have obtained virus particles from the swabs of an infected patient and the viral RNA has been isolated. In the first step, the RNA is reverse-transcribed into DNA. For this, one needs free nucleotides, an enzyme called reverse-transcriptase, and a so-called primer. The nucleotides are the raw material from which the DNA is assembled. The reverse-transcriptase is the enzyme that reverse transcribes the RNA into DNA. However, it needs a starting point, and this is where the primer is important. The primer is a short DNA fragment that matches the beginning of the RNA strand we want to reverse transcribe. Once the primer binds to the RNA, the reverse-transcriptase starts adding one nucleotide after the other to the primer, synthesising the new DNA strand. Afterwards, heat (or an enzyme) can be used to separate the RNA and DNA strands. The DNA obtained in this process is called single-stranded complementary DNA (single-stranded cDNA). Now, the actual process – the polymerase chain reaction – begins. The goal is to multiply the cDNA to obtain millions of copies over time.

For this, we again need free nucleotides as raw material, primers, and an enzyme, the Taq polymerase. The Taq polymerase is the enzyme that replicates the DNA, and it also requires primers as starting points. One of the primers is the same one as for the reverse-transcription. The other one matches the other end of the fragment that we want to replicate. The tricky part is designing these primers. We need to know the sequence of the viral RNA to be able to generate matching DNA strands. Moreover, the primers need to be specific to the viral RNA that we want to detect. As before, the primers bind to the DNA and the DNA polymerase starts adding one nucleotide after the other, replicating the DNA strand to which it is attached. Once we have double-stranded DNA sequences, we can use heat to separate the two strands again (‘denaturation’). Once separated, the Taq polymerase can create two new copies of double-stranded DNA, and the replication cycle re-starts with twice as many DNA copies as before. In this way, over the course of many replication cycles, the number of DNA copies increases exponentially. But of course, we only want this to happen if the patient swab really contains the SARS-CoV-2 virus. This is why the primers need to be specific to SARS-CoV-2. Otherwise, other RNA or DNA would be multiplied in the same way.

Fig. 3: Schematic of the steps of RT-PCR.
Real-time RT-PCR is a standard laboratory technique. However, it needs to be tailored to the specific RNA that we want to detect. The first ones to do this for the novel coronavirus SARS-CoV-2 were researchers from the Charité Berlin (Victor Corman and Christian Drosten) along with collaborators elsewhere in Europe and in Hong Kong.

The test itself takes a few hours to complete. The problem is that it a priori requires both appropriately equipped laboratories and that swabs to be shipped to these labs. To scale up testing and to obtain results more rapidly, tests that can be performed locally in hospitals or even at home are favourable. Some companies have developed devices that perform the described RT-PCR tests within a single box; one just needs to insert the swab. These can be used in any hospital to perform what is called point-of-care testing, e.g. testing close to the patient and not in some far-off laboratory.

Antibody tests

Another way to test for a viral infection is to screen the patient’s blood for antibodies to the virus. Such tests are becoming increasingly available for SARS-CoV-2. All antibody tests rely on the same principle. A blood sample (or plasma or serum) is taken from the patient and exposed to the viral antigen. If the patient’s blood contains the right antibodies for the virus, they will bind to the antigens. These bound complexes are then made visible in some way (see Box on rapid antibody tests for an example). Some of these tests are performed in labs but there are also rapid tests that can be done anywhere. They look very much like pregnancy tests and are quick and easy to perform: they require just a few
drops of blood, and give a result within 20 minutes. It is thus possible to screen a very large number of people for the antibodies.

These tests are not a replacement for RT-PCR tests, however. Our bodies require some time to generate the antibodies, meaning that the antibodies are not detectable at the beginning of the infection. RT-PCR tests, on the other hand, are extremely sensitive and can detect the virus very early on.

Human antibodies

Our body builds five different kinds of antibodies, which are called IgM, IgG, IgA, IgD, and IgE ('Ig' stands for 'Immunoglobulin', which is another name for antibodies). They are all specific to the same viral antigens but they differ in their molecular structure and serve different functions during the immune response, e.g. by penetrating into different tissues or by attacking pathogens in different ways. The first batch of antibodies is released even before the B-cell line has undergone somatic recombination and is hence not yet highly specific to the antigen. The first antibody to be built and released by B-cells upon activation is IgM. IgM has 10 binding sites at which it can attach to the viral antigens. It therefore binds strongly despite moderate specificity and is the first line of defense of the adaptive immune system. It can be detected quite early in the infection. The second antibody that is very important in the immune response to SARS-CoV-2 is IgG, which is much smaller than IgM and can be transported from blood into tissues. It can be detected only at later stages of the infection. Antibody tests mainly look for these two antibodies.

However, they serve a very important purpose. Antibodies are still detectable months after infection, long after the virus has been cleared out. Thus antibody tests allow us to find out in hindsight who was infected (remember that many people only develop mild symptoms or none at all!). This is very important information both for understanding how the disease spreads and potentially for future planning. For example, if we know what percentage of a population was infected, we know how much the virus has spread undetected. Moreover, antibodies are an indicator of immunity to future infection. While we do not know how long immunity to SARS-CoV-2 lasts, we assume that patients remain immune at least for a few months, maybe years (but there is also concern that immunity may actually wane quickly). People who are immune can safely return to work. A lot of research is currently focused on finding out more about immunity to SARS-CoV-2 and how reliably it is predicted by the presence of antibodies. (The relation between the presence of antibodies and immunity is unfortunately not clear-cut, e.g. immunity also depends on the amount of antibodies present and other factors.)

Caveats of testing

There is a very important caveat to testing, which is that no test is perfect. For example, antibody tests can fail to detect antibodies that are actually present (‘false negative re-
result’). Likewise, they can erroneously detect antibodies even though the patient does not have them (‘false positive result’). This can, for example, happen if the test cannot perfectly distinguish between antibodies to SARS-CoV-2 and antibodies to the common cold coronaviruses. For the Cellex antibody test described in the box, the probability of a false negative result is given as 6.2%. The probability of a false positive result is 4%. What does this mean?

### Rapid antibody tests

As an example of an antibody test, we here describe the setup of the test developed by the company Cellex™. It requires blood obtained by venipuncture (or plasma or serum) and is evaluated in a lab. The blood sample is placed in a well that contains the viral antigens, where the antigens have been prepared to have gold nanoparticles attached to them. If antibodies are present in the blood, they will bind the antigens. The well also contains rabbit antibodies, which are structurally different from the human ones. Again, gold nanoparticles are attached to them. This liquid mixture then migrates along a membrane through capillary action. The membrane contains two test lines for antibodies – one looking for IgM and one looking for IgG – and a control line. The test lines consist of so-called anti-human antibodies. These are molecules that bind the human antibodies.

![Fig. 5: Schematic of a rapid antibody test.](image)

When the liquid flows by, the antibodies – if present – stick to the corresponding test line. As these antibodies carry the viral antigens with the gold particles, the gold particles accumulate at this line, and a red stripe appears. If no antibodies are in the blood, the line remains invisible. The control line consists of anti-rabbit antibodies that bind the rabbit antibodies such that the control line turns red as well. If it does not turn red, we know that the liquid has not properly migrated along the membrane and thus we cannot trust the test results.

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not working</td>
<td>No antibodies detected</td>
</tr>
<tr>
<td>Working, no antibodies detected</td>
<td>IgM detected</td>
</tr>
<tr>
<td>Working, IgM detected</td>
<td>IgM and IgG detected</td>
</tr>
</tbody>
</table>

![Fig. 6: Attachment of the antibodies to the test and control lines.](image)

![Fig. 7: Some possible test results and their interpretation.](image)
Imagine we test 10,000 people of whom 1% have antibodies to the virus. From the 100 persons who have been infected, 94 will be correctly identified as immune by the test (assuming here for simplicity that having antibodies is equivalent to being immune, but remember that it is more complicated in reality). 6 will be erroneously told that they are not immune – they receive false negatives. More problematically, 396 out of the 9900 people who do not have antibodies will get a false positive – their test results will say that they have immunity to the disease even though they do not. This means that of 396 + 94 = 490 people with a positive result, only 94 are actually immune. In other words, the probability that a person who received a positive test result is really immune is only 94/490 ≈ 19%. However, in groups with a higher fraction of immune individuals, the test gives more reliable results. E.g., if 20% are immune, the probability that a positive test result is correct is about 85%. Moreover, by testing the same individuals repeatedly, we can reduce the error rate. Let us return to the example with 1% immune people. If we test an individual twice and he/she receives a positive result in both rounds, the probability that the person is really immune is ≈ 85% (assuming that the test results are independent from each other, which does not need to be the case). If we test three times and all three tests are positive, the person has the antibodies with 99% certainty.

Overall, we see that these tests are not suitable to screen entire populations since the antibody prevalence is still low among the general public. However, for specific scenarios they can be extremely useful, for example in cohort studies with repeated testing.

Conclusion

For now, no one knows what the next months will look like and when and how we will transition back to normal life. But knowledge about the new virus is being generated at incredible speed. All over the world, scientists are working to learn more about its biology, to find cures, and to develop a vaccine. Others try to assess the effectiveness of control measures and to identify the most promising strategies for controlling the epidemic while easing or ending the strict confinement rules. These tremendous efforts are critical to defeating the virus and returning to our normal lives.

Other useful resources

- A collection of material on the Corona crisis: http://web.evolbio.mpg.de/evoltheo_corona
- Video by Vaughn Cooper (University of Pittsburgh) on “The ongoing evolution of SARS-CoV-2”: www.youtube.com/watch?v=7gCY9tP981I&feature=youtu.be
- Spread of SARS-CoV-2 around the world: https://nextstrain.org/ncov
- Video produced by Olivia Pham about the spread of SARS-CoV-2 based on the nextstrain report (see previous link): www.youtube.com/watch?v=_re0HORRLZ8
- Video by Pleuni Pennings (SFSU) and colleagues on “How does SARS-CoV-2 spread?”: https://abetterscientist.wordpress.com/2020/03/28/new-video-about-how-sars-cov2-spreads
- Text by Trevor Bedford about cryptic transmission of SARS-CoV-2 in Seattle: https://bedford.io/blog/ncov-cryptic-transmission
Acknowledgements

I thank Barbora Trubenová for making the figures on RT-PCR, real-time RT-PCR, and antibody tests, Jenna Gallie for explanations on the effects of Taq polymerase on the reporter probe, Sally Otto, Ailene MacPherson, Barbora Trubenová, and Kristína Hudáková for many helpful comments, and Yvonne Kemper for copy editing the article.

Researchers find out more about SARS-CoV-2 every day. This article is a snapshot. As knowledge grows, we may find out that some matter is different from what we previously thought, and the information provided here may be outdated.