Student's Guide Contents

Background and significance

Scenario overview

As a healthcare provider, your job is to diagnose your patients and provide the best treatment possible based on that diagnosis. Sometimes, however, illnesses can be difficult to discern. Symptoms for many diseases are similar and untangling whether a patient has a run-of-the-mill cold, seasonal allergies, or a dangerous virus is not always easy. To complicate things even more, new viruses entering the human population are always a potential threat. Now, four patients have just arrived at your clinic complaining of flu-like symptoms. Are they suffering from a typical case of seasonal flu, the new and dangerous novel influenza Q virus (or nIQV – pronounced "nick vee")[1] , or something else entirely?

Previously limited to infecting cattle and other livestock, nIQV caused mild respiratory symptoms in animals. In humans, however, nIQV has wreaked havoc; though the disease has only just begun to spread, scientists have estimated that up to 40% of patients will require hospitalization and for 4% to 8% of patients, the disease will be fatal.

Today, four patients have come to your clinic complaining of flu-like symptoms. As their healthcare provider, you'll use molecular techniques to determine which viruses have infected your patients and arrive at conclusive diagnoses.

What are viruses?

Many common illnesses, including the flu and the common cold, are caused by viruses. Viruses are infectious agents with simple structures; at their most basic, they are little more than a protein coat wrapped around a small genome made of RNA or DNA (Figure 1). Because they lack organelles and other cellular machinery, they are unable to do most of the things a typical cell does—including reproduce independently of a host. Because of this, viruses are generally considered nonliving.

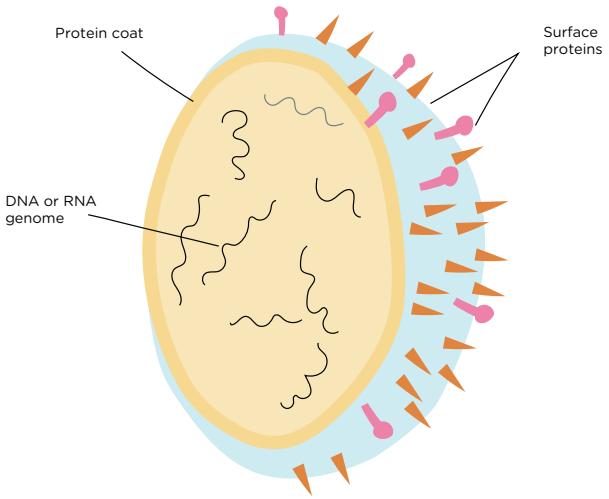


Figure 1 Structure of a virus

[1] nIQV is a fictional virus, but its story, as outlined here, represents how real diseases may emerge and spread through the human population.

To reproduce, a virus must invade a living cell and use its machinery to make copies of itself. It does this by physically attaching to and emptying its genetic material into a host cell. How exactly it does this varies from virus to virus, but the end result is the same: with the viral genome unleashed, the host cell begins to manufacture viral proteins and replicate the viral DNA or RNA. Newly made proteins and copies of the viral genome then come together to form an army of new virus particles (Figure 2). These viruses emerge from the host cell, moving on to infect new cells—and new hosts, if they're able to find their way out of the body, say though a cough or sneeze—and the cycle repeats itself.

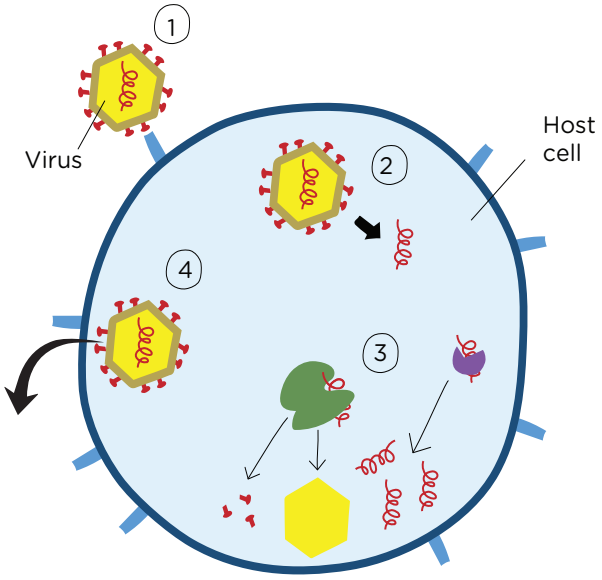


Figure 2 Viral replication cycle. 1 – Virus particle binds to and enters host cell. 2 – Virus particle sheds its coat, releasing its genome into the host. 3 – Host cell copies viral genome and expresses viral proteins, which come together to make new virus particles. 4 – Newly assembled virus particles emerge from cell, ready to infect new host cells.

A single virus particle can turn into thousands very quickly. Sometimes, the newly manufactured virus particles are packaged into capsules and released from the cell in a slow trickle. In other cases, though, replication is so fast and relentless, it continues until the host cell literally explodes, releasing a swarm of newborn virus particles. Regardless of how they exit the host cell, each of these newly released virus particles has a chance to infect new cells and, in turn, make thousands of copies of itself. In this way, the growth of the viral population is exponential. While an infection can be disastrous for the host, the speed and efficiency of viral replication represents a success from the point of view of the virus. Despite their small size, simple construction, and total dependence on living cells, viruses are remarkably efficient in achieving the primary functions of all biological entities: to persist and replicate.

The shape-shifting virus

Viruses are typically limited in the range of hosts they are able to infect. Viruses use dedicated proteins on the surface of their coats to recognize host cells. These surface proteins attach to a specific host cell or cells by binding to receptor molecules that stick out from the host cell's membrane. Receptors are normally used by host cells for cell-to-cell communication, sensing the environment, and adhering to other cells, but viruses hijack these structures for their own purposes. For this reason, a virus can only infect a given cell if that cell displays specific receptors on its surface (Figure 3). This is why you can't typically catch a cold from your dog—the receptors on your cells and a dog's cells are different enough that viruses can only recognize one or the other.

If the surface proteins on a virus change, however, they may gain the ability to bind to slightly different types of receptors, and therefore may be able to infect new hosts. This can happen as the result of a *mutation*. Viral genomes, like the genomes of living things, are subject to occasional errors in DNA or RNA copying that may be passed down to subsequent generations. While many of these mutations make it harder for a virus to replicate, and others have no effect on how a virus behaves, some may confer traits that help a virus infect a new species. Mutations in the genes encoding viral surface proteins may enable viruses to bind to host cell receptors they could not recognize before. Thus, through random mutations, the virus may gain the ability to infect a new species that it was previously unable to infect. A mutation event like this is thought to have triggered the COVID-19 pandemic; scientists believe a mutation in a strain of coronavirus carried by bats allowed the virus to jump species and eventually infect human cells. The subsequent spread of this virus from human to human led to a wave of disease that infected millions across the world.

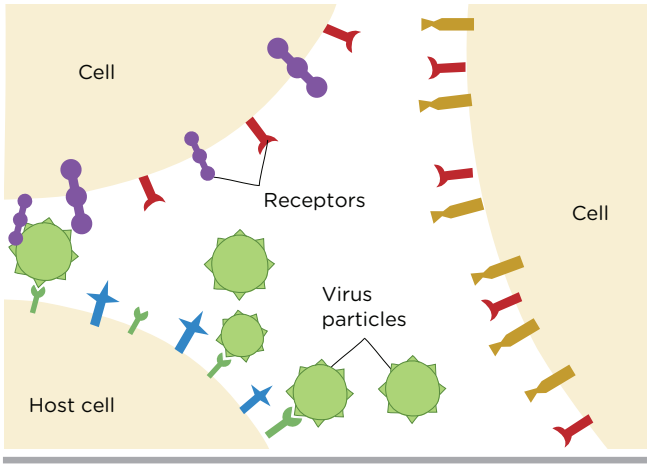


Figure 3 Virus particles can only enter cells expressing specific receptors

You may wonder why we rarely, or never, witness the emergence of a new living animal or plant species in our lifetime, yet the past few decades alone have seen the emergence of Zika, MERS, swine flu, and the COVID-19 pandemic—all of which are viral illnesses. While animal evolution tends to happen on long timescales well beyond the scope of a single human lifetime, viral evolution can proceed so quickly that it can transform the way entire human societies live.

There are three main reasons why viruses evolve so rapidly. The first is their abundance. Scientists have estimated that 10^{31} virus particles exist on Earth. That means viruses outnumber even a very abundant population of organisms—say, insects—by a factor of 1 trillion. With so many individual virus particles, there are that many more opportunities mutations to occur. The second reason is that, as we've seen, viruses reproduce prolifically. A single virus particle typically makes hundreds to thousands of copies of itself upon infecting a cell; each of its "offspring" can in turn multiply itself hundreds to thousands of times. This scale of replication offers the chance for mutations to spread widely in just a single generation. Compare this to human reproduction, where a mutation in one individual is spread to just a handful of individuals over one generation. The third reason is that in making copies of themselves, viruses make frequent errors. This is especially true for viruses with an RNA-based genome. RNA replication is dependent on a different enzyme than the one that guides DNA replication—one that is more likely to make errors while copying. So even if viral populations were similar in size to animal and plant populations, the likelihood of mutation would still be greater among viruses.

Identifying viral infections

Viruses cause different symptoms depending on what body systems they infect. Even within the body of an infected individual, the fact that different types of receptors are expressed by different cell types means a virus will invade only certain tissues. Some viruses, like the influenza viruses that cause seasonal flu, infect cells in the respiratory tract. As they take over the cells that line our airways, they cause respiratory symptoms like sneezing and coughing. The fact that these symptoms promote the spread of the virus from person to person may be no coincidence, since sneezing and coughing help the virus reach new hosts they can infect.

Viruses that infect the same body systems can produce very similar symptoms, making it difficult to tell what virus underlies a disease. This is particularly true for respiratory viruses. A cold and a case of the flu may appear quite similar, both causing a cough, nasal congestion, and body aches. In fact, the term "a cold" is a catch-all term that describes several different types of infections—both viral and bacterial—all with similar symptoms and severities.

Doctors often use laboratory tests to determine the identity of the pathogen underlying a disease. Among the most common viral tests are nucleic acid detection tests. You will be using this method to test patient samples today. Nucleic acid tests look for the presence of viral DNA or RNA in a sample, taking advantage of unique genetic sequences present in the virus. If we find a viral sequence in a sample from a human patient, we can presume the patient is infected with that virus.

A nucleic acid detection test typically involves the following four steps:

- **Step 1: Collect patient sample.** Technicians must sample patient tissue that the virus of interest will have infected. For a virus like seasonal influenza, a technician might swab the back of the throat to collect some of the respiratory cells the virus is equipped to invade.

- **Step 2: Extract genetic material.** To isolate nucleic acids from the surrounding biological material, cells are ruptured to release their DNA and RNA, and lipids and proteins are filtered out of the sample. The DNA and RNA is also stabilized during this step so it isn't broken down by subsequent processing.

- **Step 3: Amplify viral sequence.** In this step, a gene segment specific to the virus of interest is *amplified,* or repeatedly copied, so it is much more abundant than the genetic material we aren't interested in (*e.g.,* the patient's own DNA and RNA, which is also present in the sample). Most commonly, the technique polymerase chain reaction (PCR) is used for amplification; because of this, nucleic acid detection tests are sometimes referred to as PCR tests. Because PCR can only be used to amplify DNA and not RNA, detection of RNA viruses requires an additional *reverse transcription*

(RT) step where RNA is converted to DNA before PCR. The process of reverse transcription followed by PCR is called RT-PCR.

- **Step 4: Visualize results.** Finally, the test result is ready to be read. The amplified DNA is visualized to determine whether the virus of interest was present in the patient sample. A patient is said to "test positive" for a virus when we see amplified viral DNA in their sample. If we don't see amplified viral DNA, we would say they have "tested negative" and conclude that they are not infected. Often, a machine is used to detect amplified DNA in a process called quantitative PCR (qPCR). Today, however, we will directly visualize amplified PCR product ourselves using gel electrophoresis.

Today you'll experience how fast, precise, and powerful nucleic acid tests can be. In recent decades, development of nucleic acid detection methods has expanded our ability to rapidly and specifically detect infectious agents, allowing us to combat infectious disease, inform patient treatment, and better understand the processes by which viral pathogens emerge and spread.

Today's lab

In today's lab, you will use a nucleic acid detection test on samples taken from four patients displaying flu-like symptoms. Your patients may be experiencing ordinary seasonal flu or they may have been infected with the novel influenza Q virus (nIQV), an emergent RNA virus that, through a mutation, recently jumped from cattle into the human population.

With a fatality rate of 4% to 8%, nIQV is much deadlier than the seasonal flu, which kills roughly 0.1% of infected individuals. But despite this difference in severity, early nIQV symptoms look remarkably like the seasonal flu. Both viruses cause fever, aches and pains, fatigue, and a cough. Occasionally, both viruses cause nasal congestion, leading some sufferers to believe they are experiencing seasonal allergies or a particularly nasty case of the common cold. Your task will be to use molecular techniques to get your patients clear answers on the nature of their infections.

The patient samples you will be given were prepared as follows. A technician first took a nasal swab to collect tissue from the patient's airways. RNA was extracted from the sample and RT-PCR was used to convert this RNA to DNA, then amplify viral sequences from the samples (Figure 4).
For every sample, two viral sequences were amplified:

1. A 250-base pair portion of the **seasonal influenza virus gene** for matrix protein M.
2. A 400-base pair portion of the **nIQV gene** for the surface protein HEF.

If either of these sequences is present in a patient sample, we can conclude the patient was infected with the corresponding virus. That is, if we find gene M in a sample, we will diagnose that patient with seasonal influenza. The presence of gene HEF would indicate a nIQV infection.

What if a patient tests negative for both nIQV and seasonal influenza? That could be great news: a clean bill of health!
But it could also mean our experiment did not work properly: perhaps our patient has nIQV, but our chemicals went bad and we were unable to extract nIQV RNA from our nasal swabs. To rule out the possibility of such

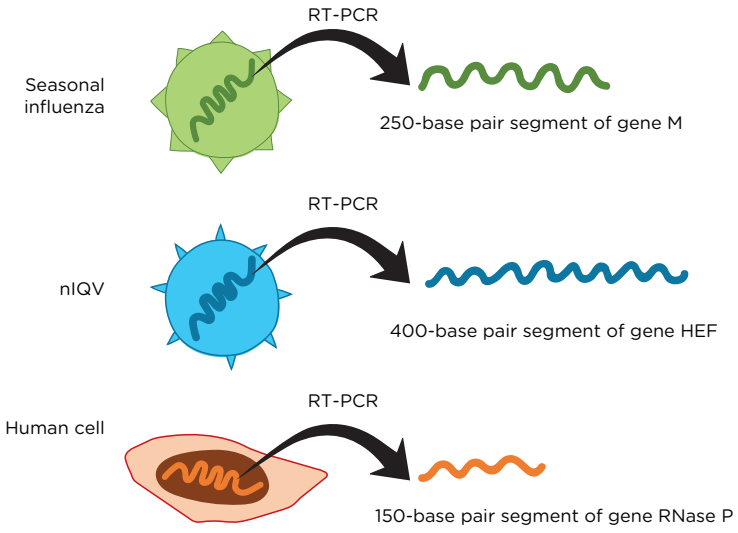


Figure 4 A schematic showing the gene sequences you will test for in this lab.

a *false negative* (a result that incorrectly indicates that an infection is not present), a third sequence was also amplified from each patient sample:

> 3. A 150-base pair segment of the **human gene for ribonuclease P** (RNase P), which is con tinually expressed by our cells, as it is involved in the process of translating RNA into protein. This sequence serves as an experimental control. It should be present in all of the patient samples—even healthy ones—since we inevitably collected some of the patients' own cells on every nasal swab. We will know our experiment has worked correctly, from sample collec tion through detection, if this sequence is successfully amplified.

What if all four of our patients test positive for nIQV? It could be that our clinic has been hit by a genuine public health disaster. Or again, it could mean our experiment did not work properly: perhaps none of our patients have nIQV, but the materials we used to collect our patients' samples were contaminated. To rule out the possibility of such a *false positive* (a result that incorrectly indicates that an infection is present), we will test a sample of human tissue known to be free from any viral agents, termed a *control* sample. If we were to find viral RNA in this sample, we would know our experiment must not be working properly, and we should re-test our patients to be sure we're getting trustworthy results.

Today, you will use gel electrophoresis to determine whether viral RNA was present in your patients' samples. Your task will be to use the data you collect to confirm your patients' diagnoses. Have they fallen ill with nIQV? The flu? Or something else?

Patient descriptions

Patient K.T.

Patient K.T. is a 66-year-old male retiree. He presents with muscle aches throughout his body and a persistent cough, with a fever of 102°F. His symptoms set in nearly a week ago, but were mild enough to ignore until today. Patient K.T. has asthma and seasonal allergies, and was hospitalized two years ago following an asthma attack.

Patient O.G.

Patient O.G. is a 37-year-old father of three. He works part-time as a nurse. He presents with wheezing and a runny nose. He does not have a fever. His symptoms began two days ago and have been steady since their onset. Patient O.G. received a flu vaccination this season and is generally in good health, aside from seasonal allergies.

Patient B.D.

Patient B.D. is a 15-year-old female who is currently a sophomore in high school. Her mother brought her to the emergency room after patient B.D. complained of being unusually tired for two consecutive days, despite sleeping 9 to 10 hours each night. Patient B.D. presents with a persistent cough and a fever of 104°F. Aside from her present illness, she is generally in good health.

Patient D.Z.

Patient D.Z. is a 42-year-old female who is visiting the United States on a family vacation from Australia. Four days into her trip, she began experiencing fatigue and shortness of breath with mild wheezing and a cough. Her symptoms have progressively worsened over the past two days, leading her to seek medical treatment today. Patient D.Z. has a fever of 102°F. She suffers from lupus, a chronic autoimmune disease, and did receive a flu shot this season, although note that the Australian flu vaccine may differ slightly from the vaccine administered in the U.S.

Predicting results

In this lab, you won't make your final diagnosis until you've run your gel electrophoresis experiment, but like any doctor, you may have hypotheses about the viruses your patients are suffering from based only on their symptoms. Consider your patients' possible diagnoses and answer the questions below. Use this chart, which outlines some of the symptoms associated with each potential diagnosis, to help you.

	Common cold	Seasonal influenza	Allergies	nIQV
Fever	Rare	Common	Never	Common
Aches & pains	Slight	Common and severe	Never	Occasional
Fatigue	Occasional	Common	Occasional	Common
Stuffy or runny nose	Common	Occasional	Common	Rare
Cough	Common	Common	Sometimes	Common

1. Patient O.G. is certain he has nIQV. As a healthcare worker at a neighboring hospital, he has cared for nIQV patients and fears he has been exposed to the virus. Based on his symptoms, do you agree or disagree with his self-assessment? Why or why not?

2. Your colleague, Nurse Li, put on a double layer of gloves and a face mask before treating patient D.Z. She is certain this patient has nIQV and is doing all she can to keep from being exposed to the virus. What evidence may have led Nurse Li to believe patient D.Z. has nIQV? What other potential diagnoses can you not rule out at this point?

3. At this point, what is your hypothesis about each of your patients' diagnoses? Of course, your diagnoses won't be final until you've received the results of your gel electrophoresis experiment.

Patient K.T.:

Reasoning:

Patient O.G.:

Reasoning:

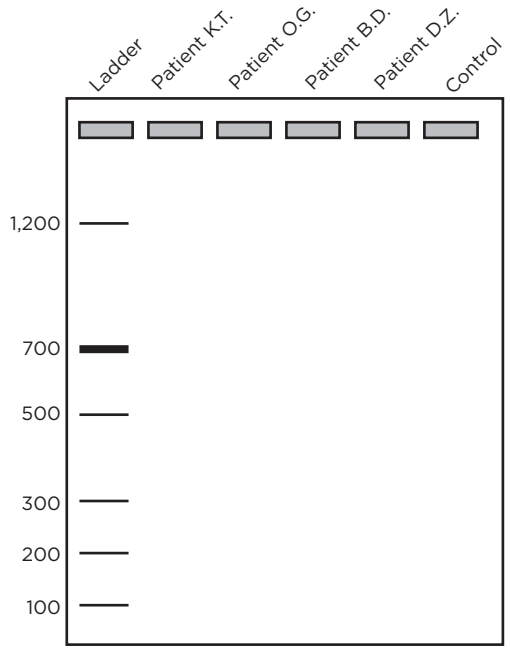
Patient B.D.:

Reasoning:

Patient D.Z.:

Reasoning:

4. The illustration to the right depicts an electrophoresis gel. Lane 1 contains a DNA ladder, showing how far bands of different size will migrate on the gel. Using the DNA segment sizes we expect from our PCR test (described on pages 14-15), predict what your gel will look like. Draw in the bands you expect to see for each patient based on your predictions from the previous question.



Laboratory guide

Protective gloves and eyewear should be worn for the entirety of this experiment.

Gel electrophoresis — Pouring gels (before or during class period)

Important Gel Pouring Notes

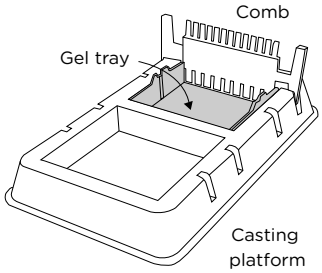
Gels can be prepared up to three days ahead of time and should be stored at ambient temperature, covered in air-tight plastic wrap and protected from light.

You will need 5 lanes plus one lane for ladder per group. If groups are sharing gels, a single lane for ladder is sufficient.

These instructions are designed for use with the blueGel™ electrophoresis system by miniPCR bio™. If using another electrophoresis system, these instructions may need to be adjusted according to the manufacturer's instructions.

1. Prepare TBE buffer (to be completed by teacher in advance)

- TBE buffer is typically provided in 20X concentration.
- Add 1 part 20X buffer to 19 parts distilled water to make 1X buffer.



2. Prepare a clean and dry casting platform with a gel tray and comb

- Place the clear gel tray in the white casting platform.
- Place a well-forming comb at the top of the gel tray.

3. Prepare a 2% agarose solution using the method indicated by your instructor

---

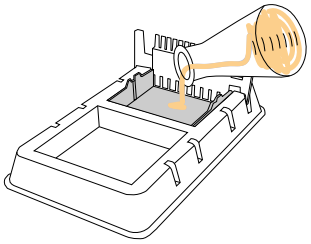
IMPORTANT NOTE: There are several ways to prepare agarose gels.

- Watch a video outlining three methods to cast agarose gels by scanning the QR code.
- Continue to pages 44-46 for detailed instructions on how to prepare agarose gels using each method.

---

4. Pour the agarose solution into the prepared casting platform with a gel tray and comb

- The agarose solution should cover the bottom of the gel tray and the bottom 3 mm of the comb (roughly the bottom 1/3 of the comb).



5. Allow gel to solidify completely and remove the comb by pulling firmly upwards

- Gels will typically be ready in about 10 minutes.
- Gel is ready when cool and firm to the touch.

Laboratory guide

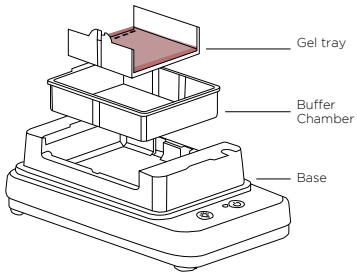
Protective gloves and eyewear should be worn for the entirety of this experiment.

Gel electrophoresis — Running the gel

These instructions are designed for use with blueGel™ electrophoresis system by miniPCR bio™. If using another electrophoresis system, these instructions may need to be adjusted according to the manufacturer's instructions.

1. Place the gel tray containing your gel in the buffer chamber

- Ensure that the clear buffer chamber is inside the blueGel™ electrophoresis system.
- The wells of the gel should be on the same side as the negative electrode, away from the power button.



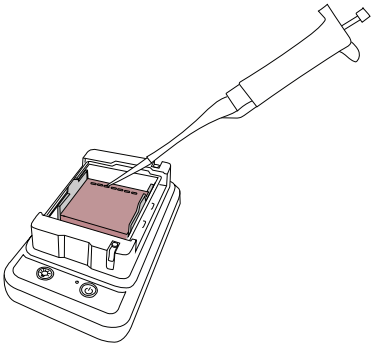
2. Add 30 ml of 1X TBE electrophoresis buffer

- The buffer should just cover the gel and wells.
- Ensure that there are no air bubbles in the wells (shake the gel gently if bubbles need to be dislodged).

3. Load samples onto the gel in the following sequence

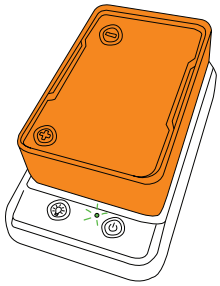
- Lane 1: 10 µL Fast DNA Ladder 1
- Lane 2: 10 µL Patient K.T.
- Lane 3: 10 µL Patient O.G.
- Lane 4: 10 µL Patient B.D.
- Lane 5: 10 µL Patient D.Z.
- Lane 6: 10 µL Control DNA

Note: Samples already contain loading dye.



4. Place the orange cover on the blueGel™ electrophoresis system

- To prevent fogging, make sure that ClearView™ spray has been evenly applied to the inside of the orange cover.
- Match the positive and negative electrode signs on the orange lid with the corresponding positive and negative signs on the blue base.
- The orange lid should sit flush with the blue base using little force.



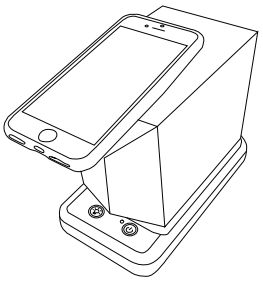
5. Press the "Run" button

- Check that the green light beside the power button remains illuminated.

6. Run the gel for 15-20 minutes

- The colored dye should progress to about half the length of the gel.
- Longer electrophoresis times will result in better size resolution.

Gel electrophoresis – Visualizing results



1. Press the "light bulb" button to turn on the blueGel™ transilluminator

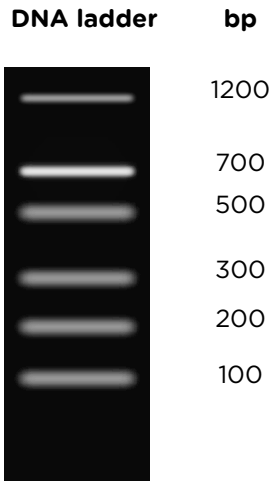
- For best viewing, dim lights or use **Fold-a-View™** photo documentation hood with a smartphone camera.
- Gels may be viewed at the end of the run or periodically throughout the run.
- If image appears hazy, wipe off the inside of the orange cover and reapply ClearView™ spray.

2. Ensure that the bands in your gel have separated enough to clearly interpret your results

- If needed, run the gel longer to increase resolution.

3. Document your results

- Compare the bands from each of your patient samples and your control to the ladder to obtain size estimates.
- Place Fold-a-View™ photo documentation hood on the blueGel™ electrophoresis system to take a picture with a smartphone or other digital camera.



Study questions - pre-lab

Review

1. Why must a virus enter a living cell in order to replicate?

2. Your friend has a cold, but insists she doesn't need to cover her mouth when she coughs. Using what you know about virus biology, explain how covering your cough helps prevent the spread of viral infections.

3. Why don't humans typically pass viruses to their pets and vice versa?

4. How was nIQV able to jump from cattle to humans? What would have had to happen to allow it to do so?

5. What are three reasons viruses mutate and change more rapidly than, say, trees do?

6. We will carry out a nucleic acid detection test on samples from each of our four patients. The samples will contain DNA and RNA from different sources: from patients' own cells, from bacteria present in the nasal passage, and potentially from viruses, too. How can we be sure that we are testing specifically for our virus(es) of interest?

7. In this lab, we use experimental controls to ensure that we don't see false positive or false negative results.

a. Explain what is meant by the term "false negative."

b. Explain what is meant by the term "false positive."

Critical thinking

8. Viruses are generally considered nonliving. What about viruses might lead someone to argue that they are actually alive?

9. Malaria is a disease caused by a parasite that invades the cells of the blood. In malaria, a single celled organism, *Plasmodium falciparum*, enters red blood cells and begins to reproduce, using the resources of the cell. Eventually the red blood cells rupture, releasing scores of new malaria parasites into the blood where they infect new blood cells. *Plasmodium falciparum* is not a virus, but in some ways it behaves like one. What are two ways malaria is like a viral infection?

10. Why do we say a video "goes viral" if it is circulated to a wide audience in a short time? Connect this to your knowledge of virus biology.

11. For each scenario below, identify whether the test result described is a false negative or a false positive. Explain your answer.

a. A nucleic acid detection test indicates a patient is not infected with nIQV when they actually are.

b. A nucleic acid detection test indicates that a patient is infected with seasonal influenza, but in fact they are healthy.

12. In your opinion, is it more dangerous to have a false negative or a false positive result? In which case are the potential consequences worse? Justify your answer.

13. In diagnosing a patient with a viral infection, why is it better to use a nucleic acid detection test than make a diagnosis based only on the patient's symptoms?

CER Table

Fill in the table based on your results from the lab. Use the rubric on the next page to help your answers.

Question:
Are any of your patients infected with nIQV?

Claim Make a clear statement that answers the above question	
Evidence Provide data from the lab that supports your claim	
Reasoning Explain clearly why the data you presented supports your claim. Include the underlying scientific principles that link your evidence to your claim	

Score	4	3	2	1
CLAIM A statement that answers the original question/ problem.	Makes a clear, accurate, and complete claim.	Makes an accurate and complete claim.	Makes an accurate but incomplete or vague claim.	Makes a claim that is inaccurate.
EVIDENCE Data from the experiment that supports the claim. Data must be relevant and sufficient to support the claim.	All of the evidence presented is relevant and sufficient to support the claim.	Provides evidence that is relevant and sufficient to support the claim. May include some non-relevant evidence.	Provides relevant but insufficient evidence to support the claim. May include some non-relevant evidence.	Only provides evidence that does not support claim.
REASONING Explain why your evidence supports your claim. This must include scientific principles/ knowledge that you have about the topic to show why the data counts as evidence.	Provides reasoning that clearly links the evidence to the claim. Relevant scientific principles are well integrated in the reasoning.	Provides reasoning that links the evidence to the claim. Relevant scientific principles are discussed.	Provides reasoning that links the evidence to the claim, but does not include relevant scientific principles or uses them incorrectly.	Provides reasoning that does not link the evidence to the claim. Does not include relevant scientific principles or uses them incorrectly.

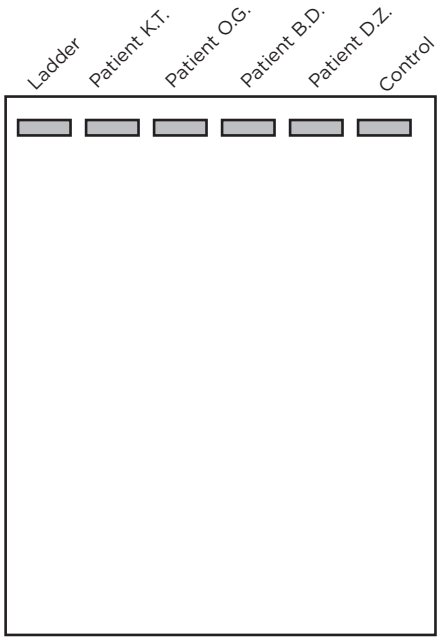
We recommend that teachers use the following scale when assessing this assignment using the rubric. Teachers should feel free to adjust this scale to their expectations.

Rubric score	3	4	5	6	7	8	9	10	11	12
Equivalent Grade	55	60	65	70	75	80	85	90	95	100

Study questions - post-lab

Interpreting results

1. The illustration to the right represents the gel you've run. Sketch in the ladder and your results for each patient. Label each band with its approximate size. Remember that in our experiment, we used PCR to amplify a 250 base pair segment of the seasonal influenza viral genome, a 400 base pair segment of the nIQV genome, and a 150 base pair segment of the RNase P control sequence.



2. After reviewing the results of your gel electrophoresis experiment, how would you now diagnose your patients?

Patient K.T.:

Patient O.G.:

Patient B.D.:

Patient D.Z.:

3. For any of your patients, is your diagnosis different from your original prediction? If so, explain how.

4. Are there any patients for whom you cannot give a definitive diagnosis? If so, identify them and list as many possible diagnoses as you can think of.

Critical thinking

5. Imagine you observed two bands in your control lane: one at 250 bp and one at 150 bp. What concern would this raise? If you were to repeat this test, what might you do differently to ensure this doesn't happen again?

6. Imagine you had four patients test positive for seasonal influenza. Your colleague insists at least some of these must be false positives; after all, what are the chances all four patients would be infected with the flu? In your control lane, you see just one band measuring 150 bp. Do you agree or disagree with your colleague's assessment?

7. What would it mean if, for one patient, you observed 3 bands: one at 150 bp, one at 250 bp, and one at 400 bp? What would you want to see in your control lane to ensure this is a reliable result?

Extension: Using Genetic Data to Track Outbreaks

Using Genetic Data to Track Outbreaks

To trace the spread of a pathogen like novel influenza Q virus (nIQV) across the globe, scientists take advantage of the fact that as viruses spread, their genetic sequences mutate. These mutations give rise to unique *variants*, or versions of the same virus with variations in their genetic sequences. By comparing the sequences of different viral samples and tracking where and when closely related variants were found, researchers can trace the geographical path of a virus through a population.

Tracking viruses requires researchers to collect patient samples just as a clinician would do to carry out nucleic acid detection tests. Researchers tracking outbreaks collect samples of infected patient tissue and extract the viral genetic material, but instead of testing for the mere presence or absence of a particular gene, they read the nucleotide sequence of the entire viral genome or vast portions of it. The more of the genome a researcher can read—or sequence—the more information they will have about the relationships between variants.

Determining how closely related two viral variants are by comparing their genetic sequences is rather simple. In short, if two variants share the same sequence at a given position in their genomes while other variants share a different one—say two guanines (Gs) where every other virus has an adenine (A) followed by a cytosine (C)—scientists make the assumption that the two viral sequences share an ancestor in which that mutation first occurred, and are more closely related to one another than other variants. While this principle is fundamentally simple, it can become complex when applied to a dataset consisting of dozens or hundreds of different sequences, each thousands of nucleotides long. Because of the sheer amount of data involved, this type of analysis is handled by computers.

```
Variant 1: AUUGGUACAC
Variant 2: AUUGGUACAC
Variant 3: AUUGGUACAC
Variant 4: AUUGGUGGAC
Variant 5: AUUGGUGGAC
Variant 6: AUUGGUACAC
```

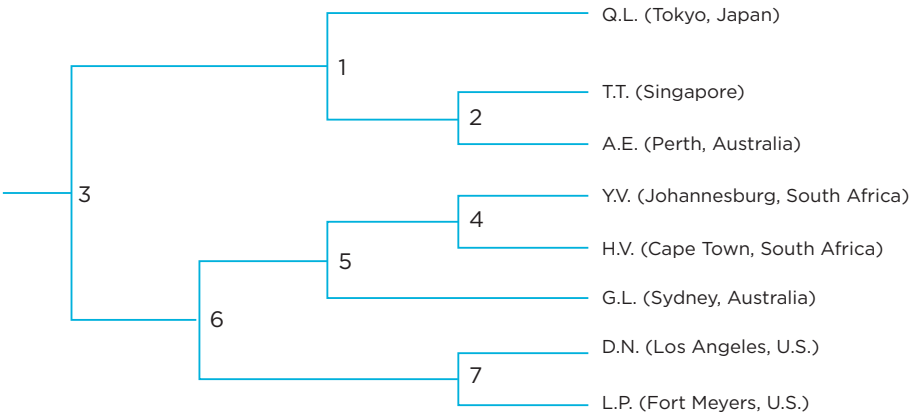
Genome sequences from 6 related viral variants. Because variants 4 and 5 share a common mutation (GG where other variants have AC), we can presume they are more closely related.

Tree thinking

Understanding how viruses are related and how they spread uses the tools of evolutionary biology and more specifically, *phylogenetics:* the study of how different species are related to each other evolutionarily. You may have experience with some of these tools; phylogenetic trees or cladograms are diagrams that depict evolutionary relationships between living things—or viruses.

Below, you can see a tree showing different nIQV virus variants. As you trace the paths of the tree from left to right, you are advancing in evolutionary time. Nodes—the points where one branch splits into two—represent the most recent common ancestor shared by the organisms on the branches. Tracing from right to left, you'll see that the sooner two organisms meet at a node, the more closely related they are.

You may see trees drawn in a variety of ways. Sometimes lines are drawn at an angle, making a series of connected "V"s. Sometimes they advance from left to right as in the tree below, but they can be oriented in any direction. Sometimes they are even drawn as a circle. There can be different reasons for drawing trees all these different ways, but universally, the most important things to look for in a tree are the nodes. Understanding at what nodes two branches connect will tell you how organisms are related to each other.



The tree above represents eight nIQV viral sequences collected from patients around the world. Each sequence bears the initials of the patient from whom the virus was isolated. In this tree the nodes have been numbered so we can easily reference them.

When reading a tree, remember that the order in which the names are written in is not important for determining relationships. What is important is determining the node at which branches meet. For example, A.E and Y.V. are written next to each other, but their branches don't meet until the deepest node in the tree (3), showing that they are only distantly related. On the other hand, Y.V. and L.P. are not close to each other the way this tree is written, but their branches meet at node 6. This tells us their viral samples are more closely related to each other than those of Y.V. and A.E.

1. From looking at the tree, which viral sequence is most closely related to the variant isolated from patient A.E.? Justify your answer by referencing a node on the tree.

2. Is the sequence isolated from G.L. more closely related to the one isolated from D.N. or H.V.? How do you know?

3. Is the sequence isolated from G.L. more closely related to either H.V. or Y.V.? Or is it equally related to both? How do you know?

4. Which node represents the most recent common ancestor of Y.V. and G.L.?

When looking at infections in a particular country, scientists will assume they all stemmed from the same outbreak if the viral sequences appear to be closely related—that is, if the number of matching nucleotides is high and the pattern of mutations is comparable. If, on the other hand, the sequences do not appear to be closely related, it is more likely that the virus has entered the country on more than one occasion.

5. South Africa, the United States, and Australia all have more than one case of nIQV. For which countries would you hypothesize that the virus entered only once and spread inside its borders? Explain your reasoning.

6. For which countries would you hypothesize that the virus entered the country more than once? Explain your reasoning.

Normally, trees like the one above would be made using hundreds or thousands of nucleotides. For simplicity, the tree above was made using 50-nucleotide segments of the nIQV genome. Sequences for each patient are listed below in the order in which they were uploaded to the database.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Q.L.	U	A	U	G	C	A	C	U	A	C	C	A	C	C	U	C	C	G	C	A	A	A	G	C	G	A	G	A	G	A	U	G	G	G	A	A	C	G	A	U	U	U	G	G	A	A	A	U	U	G
T.T.	U	A	U	G	C	A	C	U	A	G	C	C	G	C	A	A	C	G	C	C	U	A	G	C	G	A	G	A	G	A	U	G	G	G	A	A	C	G	A	U	U	U	C	G	A	A	A	U	U	G
D.N.	A	U	U	C	G	A	C	U	U	U	A	C	G	C	A	A	C	C	C	A	A	A	C	G	C	A	C	U	A	G	U	G	G	C	U	G	A	G	U	C	U	G	C	G	A	A	A	A	A	G
H.V.	A	U	U	C	G	U	C	C	U	U	A	C	G	C	A	A	C	C	C	A	A	A	C	G	C	A	G	U	A	G	U	G	G	C	U	G	C	G	U	U	G	G	C	G	G	G	A	A	C	G
A.E.	U	A	U	G	C	A	C	U	A	C	C	C	G	C	A	A	C	G	C	C	U	A	G	A	G	A	G	A	G	A	U	G	G	G	A	A	C	G	A	U	U	U	C	G	A	A	A	U	U	G
L.P.	A	U	U	C	G	A	C	U	C	C	A	C	G	C	A	A	C	C	C	A	A	A	C	G	C	A	C	U	A	G	U	G	G	C	U	G	A	G	U	C	U	G	C	G	A	A	A	A	A	G
Y.V.	A	U	U	C	G	U	C	U	C	C	A	C	G	C	A	A	C	C	C	A	A	A	C	G	C	A	G	U	A	G	U	G	G	C	U	G	C	G	U	U	G	G	C	G	G	G	A	A	G	G
G.L.	A	U	U	C	G	U	C	U	C	C	A	C	G	C	A	A	C	C	C	A	A	A	C	G	C	A	G	U	A	G	U	G	G	C	U	G	C	U	U	U	G	G	C	G	A	A	A	A	A	G
D.Z.	U	A	U	G	C	A	C	U	A	C	C	A	C	C	A	A	C	G	C	A	A	A	G	C	G	A	G	A	G	A	C	C	G	G	A	A	C	G	A	U	U	U	G	G	A	A	A	U	U	C

7. The tree shows that T.T. and A.E. are very closely related. Can you find evidence for this from their sequences? Are there any segments that they share that are not seen in the other sequences?

8. There are two major groups on the tree that arise from a split at node 3. Looking at the sequences, can you see this split? Can you identify any nucleotides or groups of nucleotides that are shared by all members of one group but no members of the other group?

In this lab, you diagnosed Patient D.Z. with nIQV. D.Z. is from Australia, but was tested in the United States. To get to the United States, she had a flight connection in Tokyo. All three of those locations have seen outbreaks of nIQV. Compare her viral sequence, which appears in the bottom row of the table above, to the sequences above it.

9. Can you say where you think D.Z. most likely caught the virus? Identify specific nucleotides that help you make your decision.

10. Add patient D.Z. to the tree. Create a new node and add a branch where you think D.Z. best fits.