Hi, I am Naomi Park, a senior staff scientist at the Wellcome Sanger Institute. I'm going to talk about the primers used for COVID amplicon sequencing. We'll talk about the primer design, the ARTIC itself, as well as other schemes, and the importance of annealing temperature in multiplex PCR and the fact it's hard on ARTIC PCR, how to improve genome coverage, and also how we've had to keep up as the pandemic has evolved, causing effects in the priming schemes.

So a the start of the pandemic, the ARTIC network quickly acted to design a multiplexed PCR that would capture the length of the 30 KB SARS-CoV-2 genome in 400 base pair chunks. They used primer scheme software for the design and openly shared the output so that teams throughout the world could use it in their own laboratories.

The 99 primer pairs are divided into two separate pools so that each sample is amplified in two separate 35 cycle PCR reactions. The diagram shows the COVID genome underneath the locations of the primers. You can see odd and even primers are distributed above and below this line. And this indicates the primer pool they go into.

So why are there two pools in the PCR? And why are they overlapping? What I mean by overlapping is the start of the left hand primer for the second primer pair being upstream of the right-hand primary pair in primer pair 1. The reason for this is that once these are amplicon sequenced, the sequence deriving from the primers themselves and not from the COVID genome-- they're from the primers.

So this area has to be independently captured by an adjacent primer pair. The reason for splitting the two pools is shown here. If they were in the same PCR reaction, the two overlapping primers would work together to form shorter products. And these would be at the detriment for the full length amplicons and therefore not target the complete genome as desired.

The speed, high quality, and open sharing of the ARTIC design led to widespread adoption. The 400 base pair insert size can be sequenced on a range of platforms, including Oxford Nanopore, and also Illumina. However, a shorter scheme was of good interest for some users of short read platforms. And a number of companies then went to design these shorter amplicon schemes, such as NEB, who released kit options, including primers for both the 400 base pair scheme but also a shorter 150 amplicon scheme. There are many other companies that have also designed other short amplicon schemes. This is just one example.

Other groups have increased the amplicon size. Those using Oxford Nanopore, it's able to see if it's much longer. So they've looked to increase the amplicon length. And here the Midnight primer sets amplifies 1,200 base pair regions of the COVID genome.

Going back to the ARTIC PCR, this slide is to demonstrate the importance of annealing temperature on the behaviour of individual primer pairs in the multiplex. PCR protocols initially released with an annealing temperature of 65 degrees for the primers to find the template.

But then frequently, a dropout of amplicon 64 was observed in the bottom of these plots. And this plot is the amplicon number sequentially from 1 up to 99 and the relative read count obtained.

Now I observe that a simple lowering of the annealing temperature by just two degrees recovered this amplicon dropout and also improved the behaviour of a number of other amplicons. So therefore we switched using 63 degrees to the PCR and shared this information widely. Now of course, the temperature in a PCR is only as good as the thermal cycle is calibrated. So this underlines the importance of regular servicing of thermal cyclists and calibration, if required.

Another optimization we have made to improve the genome coverage is to alter the relative amounts of each primer pair within the multiplex. The amount of the primer in the pool directly affects the number of reads obtained for each outcome.

So how do we do this? We do this by, first of all, making an equal volume pool of all of the primers, the same volume of each one to achieve. And we apply this to a plate of typical samples. And you see results of this in the top for ARTIC v3. There are a spread of relative primer amounts.

We then run a calibration to alter the amounts of each one to add to the primer pool, and we call this rebalancing. Now if our lab generated the same yield of products, of reads, they would cluster around 1x in this chart. Those with fewer reads are left, so they're under-covered in the left-hand of this plot. And those getting too much data than required is in the right hand.

The closer that we can get all these reads to the medium 1x coverage, the lower the total amount of sequencing is needed per sample to get sufficient coverage to use that data. And additionally, it's also less likely for sequence variation in a primer site and this variety to cause a complete dropout. It might reduce it, but still generate enough data to use.

Now in the bottom plot, you can see the results once we've applied our rebalancing to ARTIC v3. You can see far more amplicon clusters around 1x after rebalancing. And the distance between the maximum and minimum ratios are markedly reduced. So this is the rebalanced primer pool that we now use and take forward. Well, that's for v3. And we've applied the same rebalancing technique to subsequent primers.

So variation has arisen and fixed in primer bonding sequences. And these have been a challenge. So remember back to December 2021, Delta arrived and began to dominate the pandemic landscape, and with it a five base pair deletion within one of the primers for amplicon 72 in the v3 scheme broadly knocked it out. Other variants of concern, such as the Beta variant, also was causing dropouts.

So Josh Quick, and others in the ARTIC network responded and entirely redesigned the panel to improve its robustness. And during this redesign, they took all historical variants of concern and of interest into account, and v4 was born.

Similar thing happened within the arrival of Omicron in December 2021. And this caused a near knockout of amplicon 76, 79, and 90, important amplicons in the spike region. And this time, there was only redesign of affected parameters rather than them all. And this is why it's called ARTIC v4.1. So each time we have responded to redesigns by applying the rebalancing work to any modification in the primer scheme and we've had to validate these at Sanger before pursuant to operational use.

So finally, I know some monitoring primer schemes against new SARS-CoV-2 lineages. This is a great tool that's been released by NEB. And this indicates any nucleotide variants which overlap the primer scheme.

And they show this in yellow against a number of different primer schemes, either available commercially or to order them all by yourself. And this is really nice tool to enable checking potential issues. It's really important to be aware of this. And it's worth taking a look at for monitoring purposes.

Many thanks for listening.