Civet answers

1. They didn’t meet QC – one was too short and one had too high an N content
2. No. They are both in separate trees to each other, and to the focal doctor (EDB010), so none of the three sequences are closely related to each other. This makes it highly unlikely that they are connected in a meaningful way, and is likely that they all acquired their infection independently.
3. We can never say definitively that person A is connected to person B with a relatively slow-evolving virus like SARS-CoV-2. Even though the sequences are identical, there have been sequences found months apart that are identical. Genomic epi for a virus like SARS-CoV-2 is much better at ruling *out* transmission (eg question 2) than proving it.

We therefore look at all the evidence available and try to weigh up how likely the differing hypotheses of transmission are.

It is a reasonable hypothesis that these two cases are connected. The data in the timeline helps to support this hypothesis, as the timing the cases is compatible with transmission from one person to the other, based on what we know about the biology of SARS-CoV-2 infection.

More information that would be useful to help decide whether these cases are connected would be more metadata – was there an opportunity for transmission, so did they come into contact with each other when one of them was infectious? Essentially, what we need is more traditional epidemiological data to help provide context to the genomic data.

It would also be useful to know more where else both of these people caught the virus. If both individuals had a sequence very different to the local community, but very similar to each other, that supports the idea that one of them transmitted to the other. However, if they both have viruses that are similar to the background community, then it’s likely they could have both caught the virus separately. On this tree, we can see that the nearest sequence was in Glasgow and a SNP or two different. If that truly is the closest related sequence, it lends weight to the theory that the EDB010 and EDB005 are connected. However, most cases are not sampled, and so we could just be missing the key community samples. We need to know the genomic context that these sequences are in – this is where lineages can be useful again: if the two individuals have an unusual lineage for the area, then they are more likely to be connected.

The more sequencing data you have, the easier it is to ascertain if cases are connected, because you have more context. The same goes for epidemiological data.

Side note: if this was a pathogen that showed considerable within-host diversity, sequencing multiple times from one person would help provide more clarity as to whether the individuals were connected, as it is unlikely that two unconnected people would share many mutations. This approach has been used to study HIV transmission, and the transmission of some bacterial pathogens.

4. They are a SNP apart, so they’re pretty unlikely to be closely connected. It’s not totally impossible, but it’s very unlikely. It’s more likely they are both samples from the same, much wider community outbreak. Similar to above, more epidemiological and genomic context would help clarify the issue.

It’s usually worth reviewing IPC measures in any given outbreak to be on the safe side though.