Four features - already noticed or easy to discover

Changes in RBD can easily be explained

- Not quite. Residue (SARS coordinates) 472 picks up F in tissue culture from L increasing binding and infectivity (PMID: 18094188). In nCoV (position 486) F is fixed in this position - it's an L in RaTG13 and other bat viruses
- Of the 6 critical contact residues described, nCoV has mutations in 5/6 as compared to bats (https://jvi.asm.org/content/early/2020/01/23/JVI.00127-20). Most of these optimal for interaction with ACE2, including ones that mutated in SARS *during* the epidemic, leading to better binding and infectivity.
- Highly optimized for binding to human ACE2 receptor

BamHI site doesn't mean anything and is a small synonymous transition

Furin site + O-linked glycans more difficult

- Evolution, likely in non-bat reservoir
    - Selection can do amazing things
    - We're missing a lot of evolution and has never happened before in CoV
- Passage in either cells or animals as part of ongoing research on SARS-like bat CoVs
    - Selection for extremely rapid transmission
    - Could lead to acquisition of furin cleavage site
- Specific engineering as part of ongoing basic research (this 'trick' has been done in SARS)
    - Easy to introduce the site this way
    - BamHI and other sites could be used, however, many other ways to do it
    - For this type of research, investigators would have to be using a novel reverse genetics system not previously described, as opposed to those already available. This seems less likely.
- Data is consistent with all three but it is impossible to definitively prove any single scenario
    - Apart from the simplest scenario of somebody having introduced a novel gene/insert into a pre-existing virus backbone, it is difficult to see exactly what conclusive evidence would look like

Two different ways of origin of outbreak considered

- Introduction from animal reservoir - specific scenarios considered below.
- Accidental infection of researcher as part of ongoing research.
    - This type of research (including gain of function research on SARS-like bat CoVs) has been ongoing in Wuhan and other places (published papers)
    - Consideration for what containment would have been used - ranging from likely (BSL2) to unlikely (BSL4). We cannot answer this question

We discussed four specific features of 2019-nCoV:

1. Highly mutated receptor binding domain (RBD), including around key residues when comparing 2019-nCoV to a highly related bat SARS-like CoV (RaTG13; 96% identity).
    a. Nothing unusual. Comparisons were made between between SARS and SARS-like CoVs that were of similar divergence as nCoV and RaTG13. Similar levels of diversity were observed in the RBD, showing that this domain in general is highly variable, which is likely due to strong positive selection for receptor binding.
2. Reversion of gain-of-function site in the RBD to that seen in SARS.

a. Nothing unusual. The reversion of F (observed in RaTG13) to Y (observed in SARS and nCoV) is only a single base-pair change (A > T transversion). Given that the RBD is variable in general, this is not unusual.

3. Gain of BamHI restriction site in the 3' end of the spike protein of nCoV. Sequence upstream of site is somewhat variable and sequence after is conserved.

    a. Probably not unusual. The gain of the BamHI site in nCoV is the result of a single synonymous transition (T > C) that happens frequently in RNA viruses. The 3' sequence following this site is conserved not only between nCoV and RaTG13, but also more broadly across similar viruses. The site could be used to insert different versions of the spike protein gene into nCoV, but no specific data suggest that it is utilized as such.

4. Gain of furin cleavage site and O-linked glycans.

    a. Unclear. This is the first time an optimal furin cleavage site has been observed in a betacoronavirus and it is additionally coupled to a gain of O-linked glycans. Several different scenarios could explain how this was gained:

        i. Natural selection, plausibly in a non-bat reservoir / intermediate host.

        ii. Repeated passage of virus in tissue culture.

        iii. Specific engineering of the site.

In summary, after considering all things above, the only thing that remains perplexing about 2019-nCoV is the fact that it has a furin site with O-linked glycans in the spike protein between S1 and S2. It is impossible to distinguish whether this was gained due to e.g., evolution or passage, and the data is consistent with either scenario. Specific engineering is also a possibility [would make insertion really easy] [would require molecular work].

Below we briefly outline different scenarios for how nCoV may have originated.

1. Bioweapon. Highly unlikely and there is no data supporting this hypothesis.
2. Specific engineering. Unlikely as this would require significant amounts of work utilizing uncommon and currently unknown backbones of SARS-like bat CoVs. For this type of work, there are preexisting backbones that could have been utilized, but they clearly were not.
3. Tissue culture passage. The data is consistent with this scenario, although no specific hypothesis exists for how the furin site was gained, but could be due to passage in tissue culture. The virus could have been released via accidental infection of researcher(s).
4. Spillover from animal host. The data is consistent with this scenario, although no specific hypothesis exists for how the furin site was gained in an animal host. However, even though the furin site has not been observed in these viruses previously, virus evolution coupled with strong selective pressure (possibly in an intermediate host) would be capable of creating such a domain.

Background:

Bat coronavirus RaTG13 is the closest relative to nCoV-2019. Two recombinant bat viruses are close in some regions of the genomes. Pangolin virus?

Furin cleavage site rough notes about evolutionary origins:

Avian influenza example of natural and spontaneous evolution - get references and details.

There are two scenarios by which we could imagine the furin cleavage site could evolve.

1. As a human adaptation during the initial stages of the outbreak. The appearance of the mutation may have then triggered a second phase of rapid transmission. All current genome sequences are from this second phase and thus show limited diversity.

2. Adaptation to a non-human host prior to the jump to humans. This mutation is not seen in any bat coronavirus and is thus unlikely to be adaptive in those species.

Thoughts on 1: is it likely to spontaneously appear in a relatively short amount of time (and presumably small number of infections). It didn't happen in SARS with 8000 infections over 6 months. The link to the market would then be spurious - some doubt on that already. Prediction would be that the animal/environmental samples apparently found by China CDC would not have cleavage site.

Thoughts on 2: can we suggest a host where this cleavage site would likely be advantageous. Ferrets/polecats? Rodents - bamboo rats (don't know if they are popular in China)? Circulating in wild populations so limited prior human exposure until infected individual brought to the market.