Questions on my MBE paper (Apr. 14)

Xuhua Xia (xxia@uottawa.ca)

I. QUESTIONS RELATED TO DOGS AND CATS AS INTERMEDIATE HOSTS:

1. Do you suggest stray dogs as a possible intermediate for passing SARS-CoV-2 between bats and humans or between pangolins and humans?

First, it is not known if bats (or pangolins) harbor SARS-CoV-2. The bat-derived RaTG13 is too different from SARS-CoV-2. Furthermore RaTG13 is a singleton, so it is possible that the bat carrying it might get it from some other animals, and the virus had only a transient existence. The pangolin-derived high-sequencing coverage virus (pangolin|Guangdong|1, hereafter referred to as pangolinCoV) is even more different from SARS-CoV-2. Not only is pangolinCoV RNA genome has only ~90% sequence similarity to SARS-CoV-2, its "phenotype" of the genome, here represented by genomic GC content and index of CpG dinucleotide frequencies, also sets it apart (Fig. 1). GC content and I_{CpG} are typically positively correlated, and the points will scatter along the axis from lower left to upper right, but rarely go sideways (although the absolute difference is quite small).

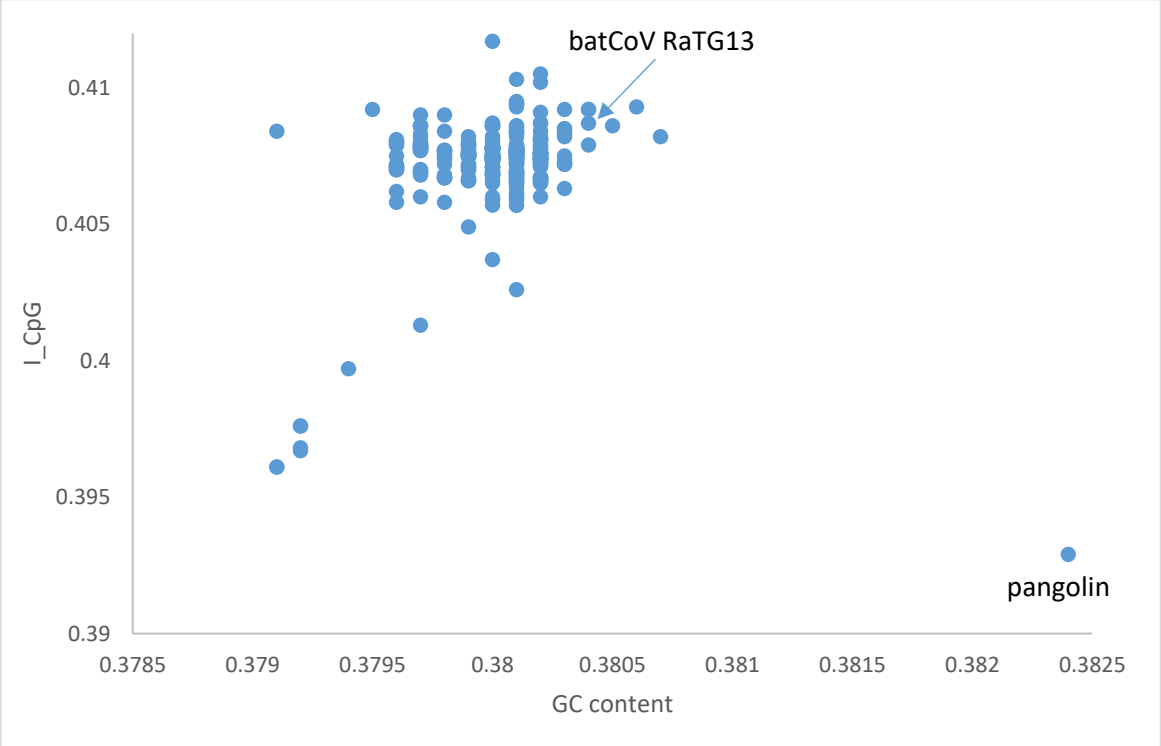

Fig. 1. A plot of I_{CpG} on genomic GC content for 437 fully-resolved (no ambiguous codes), and full-length (> 29000 nt) SARS-CoV-2 genomes with high sequencing coverage, together with bat-derived RaTG13 and pangolin-derived Pangolin|Guangdong|1 (labelled 'pangolin').

Second, what I suggested is the possibility that mammalian digestive system in general (and canid intestine in particular) might have contributed to the evolution of a SARS-CoV-2 progenitor. However, I have explained fully what brought me to this suggestion in the paper, and will provide a brief summary here.

The paper first tested the hypothesis that SARS-CoV-2 evaded human ZAP protein (zinc finger antiviral protein) by evolving reduced genomic CpG. ZAP initiates attack on viral RNA by targeting CpG in viral genomes. SARS-CoV-2, together with RaTG13 and pangolinCoV have the lowest CpG among all sequenced Betacoronavirus genomes. This is significant given previous findings that experimentally increasing CpG in the genome in a number of RNA viruses consistently reduces virulence, and decreasing CpG does the opposite. In particular, this CpG-associated decrease and increase of virulence of the virus against its host is age-related in mice.

SARS-CoV-2, TG13 and pangolinCoV are so far the most closely related among known coronavirus genomes. The simplest explanation for the low CpG among the three is that their ancestor acquired a low-CpG genome. If that is the case, then the contribution of the animal, canids or others, to the low-CpG genome would be quite early because the common ancestor of the three most likely lived more than 100 years ago.

An alternative explanation is that the common ancestor of the three, or the common ancestor of SARS-CoV-2 and TG13, did not evolve a low-CpG genome. The bat carrying RaTG13 may have acquired TG13 from other sources. This suggestion is partly due to RaTG13 being a singleton. If there are bat-derived CoV showing a gradient of CpG levels all the way down to that of RaTG13, then we may reasonably conclude that the low genomic CpG in coronaviruses can arise within certain bat lineages. However, no other bat coronavirus genome exhibit the extreme low CpG observed in SARS-CoV-2, TG13 and pangolinCoV. It is also possible that TG13 and pangolinCoV genomes might become CpG-deficient AFTER the host died. We know that pangolinCoV was isolated from pangolin carcass. We do not know if TG13, isolated from a bat in 2013 but sequence only in 2020, might also be from a bat carcass. At the time of writing my paper, the pangolinCoV genome is not available, so I checked if bats would regularly yield a coronavirus with a low CpG. TG13 is the only one with CpG as low as SARS-CoV-2.

Can a low-CpG coronavirus genome arise from another mammalian species? If we find a low-CpG genome in tissue X of animal A, then we know at least that tissue X offers a cellular environment favoring the origin of a low-CpG genome. My study follows this thread and checked all known coronavirus genomes. Surprisingly, no Betacoronavirus (to which SARS-CoV-2, TG13 and pangolinCoV belong) features a genome with CpG as low as SARS-CoV-2, TG13 and pangolinCoV. In the more distantly related Alphacoronavirus, only those infecting canine digestive system feature a low-CpG genome (and this group of canine coronaviruses include the only highly virulent canine coronavirus lineage that can spread from digestive system to other organs and cause fatal diseases in dogs). I therefore infer that the tissue X that I am looking for might be in canine digestive system. Similar tissues might also be present in other mammalian species, but they are not there in existing data. I mentioned this problem of incomplete sampling in the paper. If we have sampled all mammalian species and all their tissues, and if canine digestive system is the only tissue where low-CpG coronavirus genomes are typically found, then we can say that the tissue X exists only canids. Because we are far from sampling all mammalian species and all their tissues, we cannot conclude that a low-CpG genome can arise only from canine digestive system. I have discussed this in the paper the possibility of tissue X in other mammalian species, especially mammalian digestive systems. The emphasis in the discussion is more on the evolution of a SARS-CoV-2 in mammalian digestive system (in contrast to respiratory system) than on dogs as an intermediate species.

Finally, a cautionary note about an intermediate species. This is not a species like cats or ferrets that immediately get sick or killed by the virus (In this case the virus would expose itself immediately). What is the most dangerous is a species that carries SARS-CoV-2, but does not show obvious signs of sickness. Given that SARS-CoV-2 must have been human-

ready for quite some time and its high infectious nature, the zoonotic transmission must be a very rare event, otherwise there would be multiple outbreaks instead of a very local origin. Either the species carrying SARS-CoV-2 is very rare, or extremely few individuals in the species carry SARS-CoV-2, or the carrier species is well isolated from human populations. Thus, it is hard to find such a species. It is partly for this reason I suggested that feral dogs should be included in the virus surveillance programs.

2. A Science paper which came out last Thursday reports on a failure to infect dogs in a real experiment (cats are found to be easily infected and should be monitored, the authors are saying). Your thoughts on this?

https://science.sciencemag.org/content/early/2020/04/07/science.abb7015

I mentioned before that an animal that gets sick after SARS-CoV-2 infection is not something one should worry much about. What is particularly dangerous are those animals that get infected by the virus, carry it around and do not have obvious symptoms.

All coronaviruses from cats have relatively high CpG. This suggests that a coronavirus with a low-CpG genome has a low chance of originating in cats.

3. Could it be that there was no other host, except for humans and bats? You actually are saying that a human could be a host for a common ancestor of SARS-CoV2 and RaTG13. Wouldn't it be a more parsimonious assumption, given that no direct relatives of SARS-CoV2 or RaTG13 have been found in dogs? So, it would suggest just some gradual evolution in humans for several years, with some ZAP creating a pressure for a virus to become more virulent, correct?

Scientists do not reject the possibility of transmission directly between bats and humans, but they are not satisfied with the bat TG13 which differs in too many ways from SARS-CoV-2. Also, it is a singleton to have low CpG among all bat coronaviruses, so it is not clear if the bat might get its TG13 from some other species. Another SARS-CoV-2-like virus with a low CpG genome (and the 12 nt insertion in the S protein) would suggest a high likelihood of bats harboring SARS-CoV-2, especially if the bat carrier does not become sick. As I mentioned before, if a bat gets infected by a virus and immediately becomes fatally sick or eaten by other creatures, then it does not have a good chance of transmitting the virus to human. It is for the same reason that biologists are not satisfied with the virus isolated from pangolins.

However, one does not exclude the possibility that a virus more similar to SARS-CoV-2 may be found in pangolins or bats. SARS virus was found in bats only years after the original SARS outbreak (Although it is possible that bats/ferrets might get the virus from human, phylogenetic analysis suggests that the viral strains from bats/ferrets are more diverse and include more ancient ones, so it is more likely for us to get SARS from bats/ferrets than vice versa). It is possible that SARS-CoV-2 (and its sister lineages) might be found in bats years later as well.

The possibility that SARS-CoV-2 (or its progenitor) has been silently circulating among human populations for years is not very likely. If such a virus does exist, it should have a high chance of being isolated and sequenced after the viral outbreak, and we would find it already substantially diverged from those isolated in Wuhan but close to the root of the tree. Such viral genomes have not yet been documented. All existing SARS-CoV-2 lineages can be traced back to various lineages originally collected in Wuhan. This being said, it is possible that a SARS-CoV-2 progenitor might have infected human, did not cause serious

diseases, but then get transmitted to some other animals and rapidly evolved into the current form.

There are two main hypotheses about the origin of SARS-CoV-2. One is like the scenario of the MERS virus with multiple animal-human iterations, except that SARS-CoV-2 eventually became "fully human-ready" while MERS virus has never become quite human-ready. The other hypothesis is that the virus evolved in an animal host and went through "nearly ready" and "fully ready" stages and eventually jumped to human and immediately caused the viral outbreak. The "nearly ready" stage should have taken multiple years, but the animal host must be one that typically does not transmit viral diseases to human, so human population does not seem to have experienced infection by the virus during the "nearly ready" stage (at least the virus in a "nearly human-ready" stage has not been found yet). The existing phylogenetic analysis is in favor of the second hypothesis, with the sudden appearance of the fully-ready SARS-CoV-2 strain leading to viral outbreak in Wuhan.

4. Is human an intermediate species for pets like dogs and cats?

Whether a virus can jump from species X to species Y depends mainly on 1) whether species Y has the cellular receptor to allow viral entry into the cells, 2) whether the virus, once inside the host cell, can evade species Y's antiviral mechanisms, 3) whether the virus can output enough virions to infect others, and 4) whether the virus has gained a way to transmit from one individual of species Y to another. SARS-CoV-2 appears to use host ACE2 protein as a receptor for cellular entry, my students and I are currently studying which mammalian species have ACE2 similar to human ACE2 (sequence similarity and structural similarity).

5. You are saying that we should monitor dog's coronaviruses more closely. Is there a possibility that dogs can transmit SAR-CoV2 right now?

To transmit SARS-CoV-2 from an animal to human, the animal has to have a SARS-CoV-2 population established in a tissue that opens to the outside environment. For the time being, there is no evidence of a SARS-CoV-2 population established in a dog tissue. This could indicate that dogs are immune to SARS-CoV-2, or be due to insufficient monitoring.

6. Some cities are exterminating stray dogs now: your opinion on that?

One should test rectal swabs of stray dogs for the presence of SARS-CoV-2 before taking measures against stray dogs. If one decides to exterminate stray dogs, one not only misses an opportunity of detecting viral diversity in stray dogs, but also commits unjustified brutality against these dogs. There should be papers comparing viral biodiversity between stray dogs and pet dogs, but I was unable to find them.

7. China has excluded dogs from livestock. What do you think?

This has practical difficulty to implement. If there is an ethnic minority that runs an autonomous region where dogs are TRADITIONALLY raised and sold for meat, people in that region, being autonomous, are often entitled to keep their tradition. However, they could voluntarily give up the tradition with government subsidy to train people for other careers. People in different cultures differ in their relationship to animals. For example, some people respect horses much more than dogs. Their resistance to eat horse meat is much stronger than that to eat dog meat. I have seen tourists shocked upon learning that steak-like dish they ordered in a Swiss restaurant was horse meat.

8. Stray cats in Wuhan are found to carry SARS-CoV2. Can they infect humans?

A mammalian species such as cats that gets immediately sick upon SARS-CoV-2 infection is not dangerous, but those that carry the virus and do not become obviously sick are particularly dangerous. Whether infected cats can pass the virus to human cannot be tested experimentally. However, because infected cats can pass it to other cats and the virions passed on are the same SARS-CoV-2 virion, then it is reasonable to assume that cats can pass SARS-CoV-2 to human. I am no expert on this topic.

9. Any information on how the virus would evolve in canines?

The relevant information on how coronaviruses may evolve in canines in my paper is based on results not from SARS-CoV-2, but from canine coronaviruses (belonging to a different genus, Alphacoronavirus, in viral taxonomy). In short, one clade of canine coronaviruses (including the only highly virulent strain) have much lower CpG than all other known alphacoronaviruses. One of these low-CpG strains, when cultured in cells (which typically do not mount a strong immune response to kill the virus being cultured), quickly regained CpG dinucleotides in the genome. This suggests that some selection pushes down CpG in canine coronavirus in their natural cellular environment so CpG frequency will rebound when the selection is removed. I believe that the selection is ZAP-mediated degradation against viral RNA genome.

10. Are vaccinated domesticated pets susceptible to the virus?

If they are vaccinated with a certain coronavirus vaccine, then there might be heterologous immunity against SARS-CoV-2. If the pets are vaccinated with vaccines totally unrelated to coronavirus, then they are most likely not protected against SARS-CoV-2.

II. QUESTIONS ABOUT ZAP (ZINC FINGER ANTIVIRAL PROTEIN) AND CPG

1. What is ZAP?

ZAP is zinc finger antiviral protein (ZC3HAV1 in mammals or hZAP for human). Its structure features four fingers, with fingers 2 and 4 specialized to bind to CpG. It is a protein with about 1000 amino acids, but with different isoforms. In human it has long and short isoforms, both involved in antiviral activities but differ in function. Its isoelectric point is typically above 8, and consequently is positively charged under neutral pH. This is what one would expect for a protein that binds to RNA whose backbone is negatively charged.

2. What does ZAP do against RNA viruses?

HIV-1, Echovirus 7 and Zika virus have a low CpG genome. Experimentally increasing CpG consistently leads to reduced viral replication and virulence in the presence of ZAP protein, but such CpG-rich viruses can replicate normally in ZAP-deficient cells. This suggests that ZAP might be the only antiviral component in the host cell against CpG-rich viruses. ZAP binds to CpG and mediate viral RNA degradation. The effect is cumulative, so the viral RNA degradation may depend on how many ZAP proteins bound to the viral RNA. Human genome and consequently mRNA have low CpG because of the pervasive DNA methylation targeting CpG dinucleotides. Some DNA methyltransferases also use flanking nucleotides as part of the recognition signal, i.e., not all CpG dinucleotides are methylated the same way.

3. Where is ZAP distributed in our body?

The distribution of ZAP is quite pervasive, but especially abundant in more exposed areas. According to Human Protein Atlas (www.proteinatlas.org), ZAP is high in all parts of

digestive system (but moderate in lung and nasopharynx). It is also high in urinary bladder, testis, epididymis, seminal vesicle, vagina, fallopian tube, endometrium cervix/uterine, placenta, breast, gallbladder, pancreas, kidney, cerebellum and caudate. It is expectedly high in lymph system (lymph nodes, spleen, appendix, and tonsil). ZAP seems to be deployed to protect all vital organs. It is not clear why it is not deployed more in human lung and nasopharynx.

Dogs (both males and females) have very high ZAP mRNA in the lung, about 1.5 as high as that in spleen and more than 2 times as high as in colon from supplemental data in (Naqvi *et al.*, 2019). In contrast, human ZAP mRNA is 0.77 relative to spleen and 0.96 relative to colon (supplemental fil in Fagerberg *et al.*, 2014). If we could extrapolate mRNA to protein, then dog's lung should be well protected against SARS-CoV-2 infection. So it should be relatively rare for a dog's lung to get infected by SARS-CoV-2. However, ZAP mRNA is only moderate in dog's digestive system. In the study you mentioned above, SARS-CoV-2 was detected in dog's rectal swab, so SARS-CoV-2 may replicate in dog's digestive system as the authors acknowledged. Also, the authors of that paper inoculated the test animals intranasally and intratracheally, but did not aim to infect specifically the digestive system. To get viruses (or bacteria) to mammalian digestive system below the stomach, one probably needs to mix viruses (or bacteria) with a hearty meal that reduces the acidity in mammalian stomach fluid.

4. What ZAP environment would favor a low-CpG viral genome?

Natural selection can operate efficiently only with a large population size. If a tissue has too much ZAP to kill almost all coronaviruses (CoV), then the CoV population size would be small, and selection against CpG-rich genomic variants would not be efficient. However, if there is no ZAP, then there is no selection against CpG-rich genomic variants. So a tissue that produces a moderate amount of ZAP to select against CpG-rich genomic variants but does not reduce the viral population too much would favor viral genomes with low-CpG. Unfortunately, we do not have protein expression data across different species to make an informed inference. Human is the only species with both mRNA and protein expression data across multiple tissues, so we know protein abundance for both ACE2 (the cellular receptor that SARS-CoV-2 uses to enter our cells) and ZAP in multiple tissues. Humans have little ACE2 mRNA or protein in the lung, and moderate protein expression of ZAP in the lung. Humans have high mRNA and protein expression of ACE2 in the digestive tract (duodenum and small intestine). We also have high mRNA and protein expression of ZAP in the digestive tract (salivary gland, esophagus, stomach, duodenum, small intestine, colon, and rectum). It would be nice to have such protein data available for dogs, bats, pangolins, and others.

The observation of little ACE2 mRNA and protein in human lung, which seems to conflict with lung as a target organ for SARS-CoV-2, could have two explanations. First, there might be another cellular receptor for SARS-CoV-2 entry. Second, ACE2 is a membrane protein, and all membrane proteins are difficult to isolate and characterize. So the little ACE2 protein in human lung might be an artefact. However, others would object this second explanation by pointing out the little ACE2 mRNA in human lung, which is at least consistent with the little ACE2 protein observed in lung.

5. Is ZAP an antibody?

No. Antiviral proteins typically act before B cells can produce antibodies.

6. How much do ZAP proteins differ among different species?

This is very good questions. If a bat ZAP protein is more similar to the human ZAP than cat ZAP, then a virus that can evade the bat ZAP would be likely to also evade the human ZAP, but not the cat ZAP. This is also one of the reasons I believe that these different ZAPs may imprint different genomic signatures on the viruses.

Human ZAP has two isoforms, ZAP-L and ZAP-S (missing the C-terminal end). ZAP-L is the main antiviral component, and ZAP-S mediates the cell back to homeostasis (cellular peace. Without ZAP-S the cell will engage the virus but will be in an unnecessarily prolonged war state which can also damage the host). ZAP proteins in *Rhinolophus ferrumequinum* also have two ZAP-L and ZAP-S, but dog and cat ZAPs appear to have only ZAP-L from computational prediction. More studies are needed to characterize features and functions of ZAP proteins in different mammalian species.

7. How do viruses reduce their CpG? There is benefit for reducing CpG, but is there a cost associated with it?

Most of the SARS-CoV-2 genomes code for proteins. If a reduction of CpG leads to a nonsynonymous substitution, then there is often a fitness cost associated with it. However, SARS-CoV-2 appears to reduce CpG in three ways in coding sequences without changing amino acids. First, "NNC GNN" changes to "NNU GNN" (where a CpG has mutated to UpG). Second, ACG to ACA (where a CpG has mutated to CpA). Third, recode arginine codons CGN by alternative arginine codons AGR. These are the dominant substitutions leading to SARS-CoV-2. Nature has also created, almost miraculously, a genetic code that facilitates nonsynonymous replacement between similar amino acids: CGY(Arg) ↔ CAY(His), and CGR(Arg) ↔ AGR(Arg) ↔ AAR (Lys). As Arg, His and Lys are similar in charge and size, CGN can be channeled away from this pathways of least fitness cost.

One fitness cost of reducing CpG is almost unavoidable, and it involves codon usage. For example, in two-fold NNY codon families, human highly expressed genes typically prefer NNC codons over NNU codons. However, when "NNC GNN" changes to "NNU GNN" to reduce CpG, the C-ending codon becomes a U-ending codon. Similarly, in two-fold NNR codons, human highly expressed genes tend to prefer NNG over NNA. However, when NCG changes to NCA to reduce CpG, a major codon becomes a minor codon as a consequence.

8. How does the CpG level of SARS-CoV-2 compare to SARS-CoV and MERS-CoV?

There are 128 unique SARS virus genomes, their CpG level, measured by the I_{CpG} index, range from 0.45224 (minimum) to 0.46999 (maximum). The range is kept narrow as a consequence of mutation that tends to increase CpG and selection against CpG. There are 198 MERS virus genomes, their I_{CpG} values range from 0.55355 to 0.56598. The number of available genomes might have increased a bit during the last few months. MERS viruses have undergone multiple bat-camel-human iterations, but so far have not acquired adaptation for sustained transmission among humans. The CpG level for SARS-CoV-2 is almost uniformly 0.40.

III QUESTIONS ON MOLECULAR EVOLUTION AND PHYLOGENETICS

1. Some people have suggested that SARS-CoV-2 (or its progenitor) might have been silently transmitted in human populations for some time, "perhaps not even in Wuhan". If that is the case, then when is the beginning of this "some time"?

This "silently transmitted" SARS-CoV-2 progenitor, if exists, has been rather elusive as it has not been isolated and sequenced in spite of the extensive sequencing effort after the viral outbreak. If it were sequenced, then people mostly likely would have already identified it as an outgroup of the existing SARS-CoV-2. It is possible that this progenitor does not cause diseases so it gets no attention.

There are some practical difficulties in dating the common ancestor for SARS-CoV-2. First, the molecular clock is particularly sporadic over a short time. Some SARS-CoV-2 strains have not changed over several months, whereas some others have experienced multiple substitutions. Second, the bat TG13 and pangolinCoV are too far removed from SARS-CoV-2 lineages. When you build a tree with SARS-CoV-2 and TG3, you basically will see a branch with length of about 0.04. On one end of the branch is TG13. On the other end, you have all SARS-CoV-2 shrunk into a dot. If you magnify the dot, then you will see a tightly packed tree with short branches with average branch length of about 0.0002. Thus, standing at the point of TG13, you can't quite see which SARS-CoV-2 is closer to you. However, if you build a SARS-CoV-2 tree and grant TG13 onto different branches to generate trees with alternative rooting points, you can still reject most of the resulting trees. Third, the substitution process experienced by TG13 or pangolinCoV is likely different from that experienced by SARS-CoV-2. For example, if you estimate substitution parameters for SARS-CoV-2 sequences only, the parameter estimates likely will differ much from those when TG13 or pangolinCoV is included. However, it is possible that all these complications, hopefully, do not matter much.

2. Are evolutionary studies relevant to the fight against SARS-CoV-2?

Yes. For example, the presence of ZAP imposes selection upon the competing viral genomic variants with different CpG deficiency and consequently differential fitness. The strong CpG deficiency in SARS-CoV-2 has two implications. First, it has evolved in response to the ZAP-mediated selection. Second, its existence suggests that it has overcome the ZAP-mediated defense (which makes it dangerous). This allows one to make two inferences. First, increasing CpG in SARS-CoV-2 is likely to result in an attenuated vaccine. Second, the remaining CpG dinucleotides in the genome are likely functionally important. Manipulating this remaining CpG is likely to attenuate the virus leading to a vaccine.

Determining the origin of the virus is associated with at least three benefits. First, if SARS-CoV-2 was found in animal X, then we would distance ourselves from animal X. Second, understanding how SARS-CoV-2 progenitors evolve in animal X would help us better characterize host-parasite coevolution, trace viral evolutionary trajectory and predict the emergence of new pathogenic viruses. Third, animal X, as an intermediate or reservoir species which is typically not much affected by the virus (which is why it is difficult to find it), would help us study antiviral mechanism, characterize immune responses to the viral infection,

3. Could you clarify what you mean by the "average distance between bat CoV RaTG13 and SARS-CoV-2"?

This distance refers to evolutionary distance. It measures how two genomes have diverged due to accumulation of nucleotide substitutions. If we have 3 SARS-CoV-2 genomes (S1, S2, S3) and they diverged from TG13 by distances D1 between S1 and TG13, D2 between S2 and TG13 and D3 between S3 and TG13, then the average distance is (D1+D2+D3)/3.

4. Why does sequence divergence between bat TG13 and SARS-CoV-2 matter when considering *Rhinolophus* bats as the source of SARS-CoV-2?

Imagine a common ancestor dated in 1960s. One daughter lineage (say lineage A) from the ancestor leads to TG13, and another lineage, say lineage B yields SARS-CoV-2, probably after a number of cladogenic events. Therefore, the descending lineages in lineage B are more relevant than those in lineage A. Had we dated the common ancestor between TG13 and SARS-CoV-2 to 2018 instead of 1960s, then TG13 would be much more relevant. Many genomic substitutions could have happened since around 1960s. A closer relative would help us understand more of the origin of SARS-CoV-2. Two viruses with genomic sequence similarity >99% could reasonably be assumed to be phenotypically highly similar, but two viruses with genomic sequence similarity of 96% may not. A singleton of TG13 does not convince people that *Rhinolophus* bats are the source of SARS-CoV-2.

IV GENERAL QUESTIONS

1. What is the significance of the paper?

First, the paper tested the hypothesis that SARS-CoV-2 has evolved to evade human ZAP protein (zinc finger antiviral protein) by reducing genomic CpG. ZAP initiates attack on viral RNA by targeting CpG in viral genome, and is expected to favor genomic variants with reduced CpG. Indeed, SARS-CoV-2, together with RaTG13 and pangolinCoV have the lowest CpG among all sequenced Betacoronavirus genomes. This is significant given previous findings that experimentally increasing CpG in the genome in a number of RNA viruses consistently reduces virulence, and decreasing CpG does the opposite. In particular, this CpG-associated decrease and increase of virulence is age-related.

Second, the paper points to a new way to detect host-switching or tissue-switching events, and can improve and enhance the current viral surveillance programs. In particular, a change in CpG dinucleotide frequencies in viral genomes should be closely monitored. Because it is very rare for coronaviruses to gain a low-CpG genome, our immune system is typically not prepared/primed against it.

Third, the paper presented evidence that canids have a tissue that favors the evolution of coronaviruses with a low-CpG genome, suggesting the need for including stray dogs in virus surveillance programs.

Fourth, two implications from the study is relevant to vaccine development. First, SARS-CoV-2 is highly likely to depend on its reduced genomic CpG to evade ZAP-mediated host defense. Second, the remaining CpG in the genome are likely to be functionally important given the benefit of reducing CpG. Thus, either increasing genomic CpG in SARS-CoV-2 or modifying the remaining CpG are likely to result in an attenuated virus that may serve as a vaccine.

2. Does the paper have limitations?

Yes. As I have mentioned before, if we have data from all mammalian species and all tissues, and only canine digestive system harbors coronaviruses with a low CpG genome, then canids would be implicated as a strong candidate contributing to the evolution of low CpG in SARS-CoV-2. However, because we are far from having all data from all mammals and all mammalian tissues, the suggestion of canids as a possible host species is speculative. I have discussed this in the paper.

3. Some researchers have suggested that the virus could become less deadly over time. Others have argued that it could mutate to become more lethal. What's your take?

This is a good question. In a crowded environment when SARS-CoV-2 can easily transmit from person to person, then efficient entry into the cell, rapid viral proliferation in the infected host tissue, and consequently rapid build-up of a viral population that can go to infect others are the key contributors to viral fitness. In this case, being deadly does not reduce viral fitness. In contrast, when people practice self-isolation, a SARS-CoV-2 strain that is immediately deadly will effectively stop its own transmission and consequently reduces its fitness. In this context, viral fitness is increased by its carriers NOT exhibiting obvious symptoms to facilitate its transmission. In other words, viral fitness is context-dependent. Being cautions with self-isolation typically discourages the origin of deadly viral strains. Self-isolation has never failed in combat against viral outbreaks.

4. What is the difference between mammalian digestive and respiratory systems in terms of coronavirus evolution?

Coronaviruses infecting mammalian digestive system tend to have lower CpG in their genomes than those infecting mammalian respiratory system (However, it is difficult to perform a phylogeny-based comparison). Because SARS-CoV-2 has very low CpG, it is more likely to result from evolution in a mammalian digestive system than in a mammalian respiratory system.

5. Does your Fig. 2 and Fig. 3 include pig viruses?

Yes. Fig. 2 include many viral strains of porcine hemagglutinating encephalomyelitis virus, and Fig. 3A includes many strains of porcine epidemic diarrhea virus and swine enteric coronavirus. They all have much higher CpG dinucleotide frequencies than SARS-CoV-2.

6. Regarding the outlier in canine CoV sequences (Fig 3A), KC175339, I understand that cell culture might have relaxed antiviral selection pressure, but why would it have only increased in I_{CpG} but not that much the genomic GC proportion?

Typically, when CpG changes to UpG, not only I_{CpG} decreases, genomic GC content decreases as well. However, different viral lineages have significantly different slopes when one regresses I_{CpG} on genomic GC content. I_{CpG} increase rapidly with genomic GC content in some viral lineages, but slowly in some other viral lineages. This can be explained by the joint effect of APOBEC3G and ZAP. APOBEC3G decreases C (shifting points leftward in Fig. 3A) but has little effect on CpG (i.e., it does not shift I_{CpG} downward). ZAP selects specifically against CpG. When a cellular environment has ZAP but little APOBEC3G, then I_{CpG} will go down more steeply with genomic GC than when APOBEC3G acts alone.

V QUESTIONS THAT ARE SOMEWHAT POLITICAL

1. Why is China often a source of deadly viral pathogens? First there was SARS, and now COVID-19.

A large country with a large population will have a greater chance of having a zoonotic transmission and a subsequent viral outbreak than a small country with a small population, everything else being equal. China is very large, with a huge population and tremendous amount of biodiversity. USA is also quite large and has its fair share of viral origin and outbreaks. You don't expect many zoonotic transmission in Vatican because it is not only small but also well isolated from wildness.

Humans are part of nature, and contact with wild animals are inevitable, which creates opportunities for zoonotic transmission. It is important to know what potential pathogens are out there. If we know that a bat virus uses receptor X to gain cell entry, and if that bat receptor X is quite similar the human equivalent, then the virus can potentially gain entry to our cells if given the chance.