Corrections & amendments

Author Correction: Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins
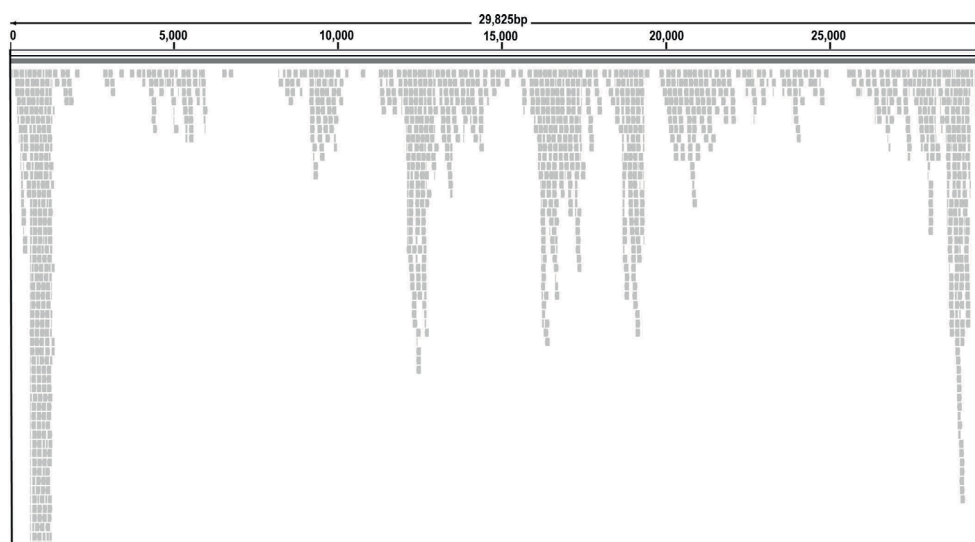
Kangpeng Xiao, Junqiong Zhai, Yaoyu Feng, Niu Zhou, Xu Zhang, Jie-Jian Zou, Na Li, Yaqiong Guo, Xiaobing Li, Xuejuan Shen, Zhipeng Zhang, Fanfan Shu, Wanyi Huang, Yu Li, Ziding Zhang, Rui-Ai Chen, Ya-Jiang Wu, Shi-Ming Peng, Mian Huang, Wei-Jun Xie, Qin-Hui Cai, Fang-Hui Hou, Wu Chen, Lihua Xiao & Yongyi Shen

Check for updates

In this Article, data in Extended Data Table 3 and Extended Data Fig. 4 were mislabelled and attributed incorrectly. The Pangolin-CoV genome reported was built using the metagenomic dataset described previously by Liu et al. in *Viruses*[1] (ref. 15 in our *Nature* paper) and targeted PCR. The *Viruses* article[1] and its metagenomic data (PRJNA573298) were cited in our Article. All 21 animals from the March 2019 seizure of smuggled pangolins were used for our Article, and 11 of these were used in the *Viruses* study.

Wu Chen, one of the corresponding authors of the *Nature* Article as well as a coauthor of the study by Liu et al., provided all the samples and the associated data for both studies. Dr Chen, who archives data from clinical specimens by animal ID, provided all samples and data for the *Nature* Article with matched IDs, including some metagenomic data. Because the numbering system for the metagenomic data was different from that used for the *Viruses* paper, the informaticians of

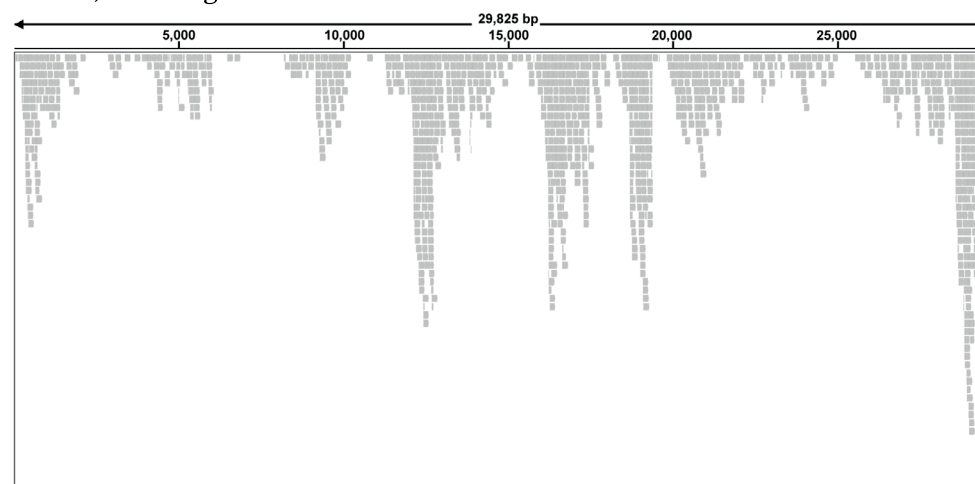Original, incorrect figure

Revised, correct figure

Fig. 1 | This figure shows the incorrect, as-published version and the corrected version of Extended Data Fig. 4 of the original Article.

our team believed that metagenomic data from M2, M3, M4 and M8 provided by Dr Chen were new. Instead, the samples labelled M2, M3, M4 and M8 in the Extended Data Table 3 of our Article correspond to samples labelled lung07, lung02, lung08 and lung11 in the *Viruses* paper. The lack of face-to-face meetings imposed by various restrictions due to COVID-19 among the four research groups involved with the study led to a delay in finding out the problem. Extended Data Table 3 of the original Article has been corrected to clarify the relationship between these samples.

The original Extended Data Table 3 listed nine animals, including the four samples (M2, M3, M4 and M8) that overlap with data from the *Viruses* paper. The corrected table lists 12 pangolins. Data from two samples (P59 and P60) submitted to NCBI during the initial submission to *Nature* were inadvertently omitted from Extended Data Table 3, and data from A22 were submitted to NCBI later. In revising our *Nature* Article, we became aware of both the overlap of metagenomic data with the *Viruses* paper and the fact that Jin-Ping Chen (corresponding author of the *Viruses* paper) was using the *Viruses* dataset in the preparation of another manuscript[2]. We therefore added new metagenomic data from pangolin A22 that were generated in March 2020 and have almost full coverage of the Pangolin-CoV genome. We neglected to upload these data to NCBI. During the preparation of another manuscript on the pangolin coronavirus[3], which used some of the metagenomic data in the *Nature* Article, we realized that data from A22 had not been submitted to NCBI. We therefore updated our dataset on 19 June 2020 and added the A22 data to BioProject PRJNA607174 (released on 22 June 2020). The labels in BioProject have been updated using the animal IDs to avoid confusion.

In addition, in Extended Data Table 3, the Chinese pangolin M10 was incorrectly labelled as a Malayan pangolin and the numbers of total reads for samples M1 and M6 were incorrect. As described in the Methods, these samples were sequenced with the paired-end approach (two reads in each pair). The original table showed values for the number of paired reads (107,267,359 and 232,433,120, respectively, for M1 and M6) instead of the number of total reads: 214,534,718 and 464,866,240. The original published Extended Data Table 3 is shown as Supplementary Information to this Amendment, for transparency to readers.

The first two sentences in the fourth paragraph misstated the numbers of samples: 'Illumina RNA sequencing was used to identify viruses in the lung from nine pangolins. Mapping sequence data to the reference SARS-CoV-2 WHCV genome identified coronavirus sequence reads in seven samples (Extended Data Table 3).' Instead these sentences should read 'Illumina RNA sequencing was used to identify viruses in the lung from **12 pangolins (including four that were reported previously**[1]). Mapping sequence data to the reference SARS-CoV-2 WHCV genome identified coronavirus sequence reads in nine samples (Extended Data Table 3)'.

The following two sentences in that paragraph said 'For one sample, higher genome coverage was obtained by remapping the total reads to the reference genome (Extended Data Fig. 4). We obtained the completed coronavirus genome (29,825bp)—which we designated pangolin-CoV—using the assembled contigs, short sequence reads and targeted PCR analysis'. This completed genome used the metagenomic dataset of lung08 (labelled M4 in the *Nature* Article) published by Liu et al.[1] in the initial metagenome assembly. We used PCR in filling the numerous sequence gaps and ambiguities in the metagenome assembly. Altogether, we obtained sequences through PCR to cover ~90% of the genome. The primers are listed in Supplementary Table 1 of this Amendment.

Extended Data Fig. 4 of the original Article was based on a composite of data from lung08 and M1. The sequence reads from metagenomic sequencing were mapped to the SARS-CoV-2 WHCV genome initially in the identification of viral sequences in the samples and in primer design. They were subsequently mapped to the

Original, incorrect table

Sample ID	Animal species	Total reads*	No. mapped
M1	Malayan pangolin	107,267,359	496
M2	Malayan pangolin	38,091,846	302
M3	Malayan pangolin	79,477,358	14
M4	Malayan pangolin	32,829,850	1,100
M5	Malayan pangolin	547,302,862	56
M6	Malayan pangolin	232,433,120	10
M8	Malayan pangolin	44,440,374	12
M10	Malayan pangolin	227,801,882	0
Z1	Chinese pangolin	444,573,526	0

* 150-bp paired-end reads

Revised, correct table

Animal ID in this study	ID in Liu et al. *Viruses* 2019	Animal species	Total reads	No. mapped
M10	/	Chinese pangolin	227,801,882	0
M6	/	Malayan pangolin	464,866,240	10
P60	/	Malayan pangolin	424,322,500	0
Z1	/	Chinese pangolin	444,573,526	0
P59	/	Malayan pangolin	393,522,404	8
M5	/	Malayan pangolin	547,302,862	56
M1	/	Malayan pangolin	214,534,718	496
A22	/	Malayan pangolin	191,038,872	5,444
M4	lung08	Malayan pangolin	32,829,850*	1,100
M2	lung07	Malayan pangolin	38,091,846*	302
M3	lung02	Malayan pangolin	79,477,358*	14
M8	lung11	Malayan pangolin	44,440,374*	12

*Source: PRJNA573298 from Liu et al. *Viruses* 2019[15]

Fig. 2 | This figure shows the incorrect, as-published version and the corrected version of Extended Data Table 3 of the original Article.

assembled Pangolin-CoV genome for confirmation of the detection of Pangolin-CoV in these samples. In both analyses, mapped reads were further extracted from the two best metagenomic datasets (lung08 and M1) at the time for additional mapping of pooled reads to guide the primer design and to check the accuracy of the genome sequence. The intention was to visually show the presence of Pangolin-CoV sequence reads in these samples, using the data from the best sample lung08 (M4 in the *Nature* Article) as an example in Extended Data Fig. 4. However, the mapping plot from pooled reads of lung08 and M1 was mistakenly used for the figure. The corrected Extended Data Fig. 4 uses data from lung08 only, and the legend clarifies that lung08 was M4 in the *Nature* Article. The original, published Extended Data Fig. 4 is shown as Supplementary Information of this Amendment, for transparency to readers.

The raw sequence data (including the trace files) generated by PCR for the assembly of the Pangolin-CoV genome have been deposited to the SRA database of NCBI (accession no. SRX9503273), the six full sequences of the S gene generated by PCR from pangolin samples in the study to GenBank (accession nos. MT799521–MT799526), and we have added a new table on primers used in PCR (Supplementary Table 1 of this Amendment).

We thank Yujia Alina Chan and Shing Hei Zhan for bringing the errors to our attention[4]. The original Article has been corrected online.

Corrections & amendments

1. Liu, P., Chen, W. & Chen, J. P. Viral metagenomics revealed sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses* **11**, 979 (2019).
2. Liu, P. et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* **16**, e1008421 (2020).
3. Li, X. et al. Pathogenicity, tissue tropism and potential vertical transmission of SARSr-CoV-2 in Malayan pangolins. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.22.164442 (2020).
4. Chan, Y. A. & Zhan, S. H. Single source of pangolin CoVs with a near identical Spike RBD to SARS-CoV-2. Preprint at *bioRxiv* https://doi.org/10.1101/2020.07.07.184374 (2020).

Supplementary information is available in the online version of this Amendment.