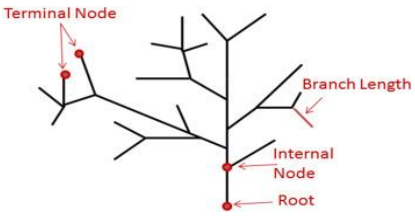
Date	Article	Sample Text	Definitions
Jan 30	**Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding**	We did next-generation sequencing of samples from bronchoalveolar lavage fluid and cultured isolates from nine inpatients, eight of whom had visited the Huanan seafood market in Wuhan. Complete and partial 2019-nCoV genome sequences were obtained from these individuals. Viral contigs were connected using **Sanger sequencing** to obtain the full-length genomes, with the terminal regions determined by rapid amplification of cDNA ends. **Phylogenetic analysis** of these 2019-nCoV genomes and those of other coronaviruses was used to determine the evolutionary history of the virus and help infer its likely origin. **Homology modelling** was done to explore the likely receptor-binding properties of the virus. After removing adapter, low-quality, and low-complexity reads, high-quality genome sequencing data were generated. Sequence reads were first filtered against the human reference genome (hg19) using **Burrows-Wheeler Alignment**.15 The remaining data were then aligned to the local nucleotide database (using Burrows-Wheeler Alignment) and non-redundant protein database (using RapSearch),16 downloaded from the US National Center for Biotechnology Information website, which contain only coronaviruses that have been published. Finally, the mapped reads were assembled with **SPAdes17** to obtain a high-quality coronavirus genome sequence.	**Sanger sequencing**, also known as the “chain termination method”, is a method for determining the nucleotide sequence of DNA. The method was developed by two time Nobel Laureate Frederick Sanger and his colleagues in 1977, hence the name the Sanger Sequence. **Phylogenetic analysis** is the means of estimating the evolutionary relationships. In molecular phylogenetic analysis, the sequence of a common gene or protein can be used to assess the evolutionary relationship of species.  **Homology modelling**, also known as comparative modelling of protein, refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"). **Burrows-Wheeler Alignment (BWA)** is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome. It consists of three algorithms: BWA-backtrack, BWA-SW and BWA-MEM. The first algorithm is designed for Illumina sequence reads up to 100bp, while the rest two for longer sequences ranged from 70bp to 1Mbp. BWA-MEM and BWA-SW share similar features such as long-read support and split alignment, but BWA-MEM, which is the latest, is generally recommended for high-quality queries as it is faster and more accurate. BWA-MEM also has better performance than BWA-backtrack for 70-100bp Illumina reads.

			SPAdes17 A chimeric read is a digital DNA sequence (i.e. a string of letters in a file that can be read as a DNA sequence) that originates from an actual chimera (i.e. an physical DNA sequence in a sample) or produced due to misreading the sample.
Feb 3	**A new coronavirus associated with human respiratory disease in China**	Phylogenetic analysis of the complete viral genome (29,903 nucleotides) revealed that the virus was most closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus Betacoronavirus, subgenus Sarbecovirus) that had previously been found in bats in China5. This outbreak highlights the ongoing ability of viral spill-over from animals to cause severe disease in humans. To investigate the possible aetiological agents associated with this disease, we collected bronchoalveolar lavage fluid (BALF) and performed **deep meta-transcriptomic sequencing.** The clinical specimen was handled in a biosafety level 3 laboratory at Shanghai Public Health Clinical Center. Total RNA was extracted from 200 µl of BALF and a meta-transcriptomic library was constructed for pair-end (150-bp reads) sequencing using an Illumina MiniSeq as previously described4,6,7,8. In total, we generated 56,565,928 sequence reads that were de novo-assembled and screened for potential aetiological agents. Of the 384,096 contigs assembled by Megahit9, the longest (30,474 nucleotides (nt)) had a high abundance and was closely related to a bat SARS-like coronavirus (CoV) isolate—bat SL-CoVZC45 (GenBank accession number MG772933)—that had previously been sampled in China, with a nucleotide identity of 89.1% (Supplementary Tables 1, 2).	Since **metatranscriptomics** focuses on what genes are expressed, it allows to understand the active functional profile of the entire microbial community. The overview of the gene expression in a given sample is obtained by capturing the total mRNA of the microbiome and by performing a whole metatranscriptomics shotgun sequencing. In genetics, shotgun sequencing is a method used for sequencing random DNA strands. It is named by analogy with the rapidly expanding, quasi-random firing pattern of a shotgun. The chain termination method of DNA sequencing ("Sanger sequencing") can only be used for short DNA strands of 100 to 1000 base pairs. Due to this size limit, longer sequences are subdivided into smaller fragments that can be sequenced separately, and these sequences are assembled to give the overall sequence.
Rec 19 Feb Pub 12 Mar	**Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Isolated in Nepal**	Sequencing was done using the Illumina MiSeq system with the Burrows-Wheeler Aligner MEM algorithm (BWA-MEM) 0.7.5a-r405 assembly method. The new genome sequence was obtained by first mapping reads to a reference SARS-CoV-2 genome using BWA-MEM 0.7.5a-r405 with **default parameters** to generate the consensus sequence. In addition, the assembly produced by MEGAHIT 1.2.9 (de novo assembly), using default parameters, was used to cross-validate with the reference-based method as an internal control. For phylogenetic analyses, SARS-CoV-2 full-genome sequences were aligned with CLUSTAL W (6) using MEGA 10.0.5. (7).	
Mar 25	**Supercomputer simulates molecular model of SARS-CoV-2**	With an estimated 200 million atoms in this model — all of whose interatomic interactions need to be calculated — this is an exceptionally daunting task. But the project's lead, Dr. Rommie Amaro of UC San Diego, has the help of Frontera, one of the world's most powerful supercomputers housed at TACC.	

		She is also not new to such feats, having done an all-atom simulation of the envelope of the influenza virus earlier this year in a paper published in ACS Central Science, which helped to identify a potentially new substrate binding mechanism.
Apr 3	**Bad News Wrapped in Protein: Inside the Coronavirus Genome**	**New York Times Article** Sources: Fan Wu et al., Nature; National Center for Biotechnology Information; Dr. David Gordon, University of California, San Francisco; Dr. Matthew B. Frieman and Dr. Stuart Weston, University of Maryland School of Medicine; Dr. Pleuni Pennings, San Francisco State University; David Haussler and Jason Fernandes, U.C. Santa Cruz Genomics Institute; Journal of Virology; Annual Review of Virology. Model sources: Coronavirus by Maria Voigt, RCSB Protein Data Bank headquartered at Rutgers University–New Brunswick; Ribosome from Heena Khatter et al., Nature; Proteins from Yang Zhang's Research Group, University of Michigan.
May	**Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods**	Twenty structures including 19 SARS-CoV-2 targets and one human target were built by homology modeling. The novel coronavirus found at the end of 2019 was named as 2019 novel coronavirus or "2019-nCoV" by the World Health Organization (WHO) on January 12, 20207,8. Since 2019-nCoV is highly homologous with SARS-CoV, it is considered a close relative of SARS-CoV. The International Virus Classification Commission (ICTV) classified 2019-nCoV as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) on February 11, 2020. At the same time, WHO named the disease caused by 2019-nCoV as COVID-19.

COVID TIMELINE

https://covidreference.com/timeline

Friday, 3 January

While examining bronchoalveolar lavage fluid collected from hospital patients between 24 and 29 December, Chinese scientists at the National Institute of Viral Disease Control and Prevention ruled out the infection with 26 common respiratory viruses, determined the genetic sequence of a novel β-genus coronaviruses (naming it '2019-nCoV') and identified three distinct strains.[2]

[2] Notes from the Field: An Outbreak of NCIP (2019-nCoV) Infection in China — Wuhan, Hubei Province, 2019−2020, China CDC Weekly, 2020, 2(5): 79-80
http://weekly.chinacdc.cn/en/article/id/e3c63ca9-dedb-4fb6-9c1c-d057adb77b57