

1 **Molecular evidence of SARS-CoV-2 in New York before the first pandemic wave**

2

3 **SUPPLEMENTARY INFORMATION**

4

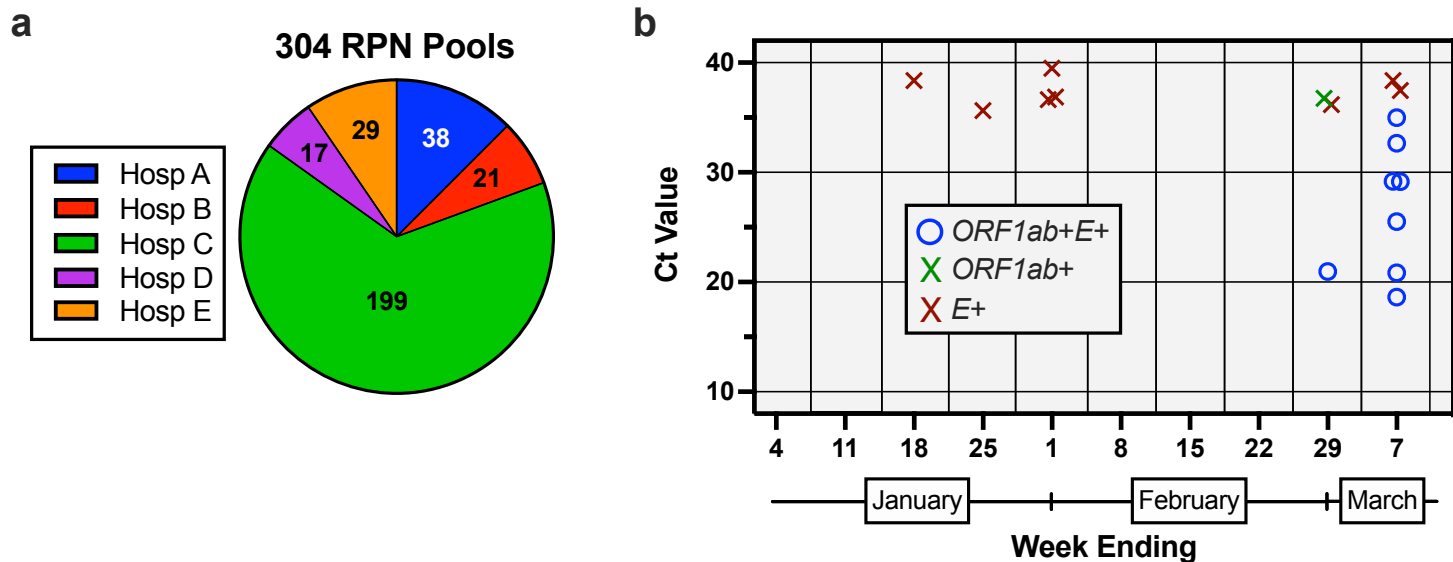
5 **Supplementary Fig. 1. Hospital distribution and SARS-CoV-2 quantitation of negative pools. a**

6 304 RPN pools were generated from nasopharyngeal swab specimens collected at distinct MSHS sites  
7 (A-E) from 30 December 2019 to 7 March 2020 (Hospital A, blue; B, red; C, green; D, purple; E,  
8 orange). **b** SARS-CoV-2 NAAT quantities in RPN pools by week. Cycle thresholds (Ct) of pools with  
9 detectable RNA by clinical diagnostic SARS-CoV-2 RT-PCR are depicted. If only the *ORF1ab* target or  
10 only the *E* gene target was detected in a pool, an X denotes the corresponding Ct value in green or red,  
11 respectively. If both targets were detected, the average of the Ct values of both detected targets is  
12 depicted by a blue circle for that pool.

13

14

# Supplementary Fig. 1



**Supplementary Fig. 1. Hospital distribution and SARS-CoV-2 quantitation of negative pools.** (A) 304 RPN pools were generated from nasopharyngeal swab specimens collected at distinct MSHS sites (A-E) from 30 December 2019 to 7 March 2020. (B) SARS-CoV-2 NAAT quantities in RPN pools by week. Cycle thresholds (Ct) of pools with detectable RNA by clinical diagnostic SARS-CoV-2 RT-PCR are depicted. If only the *ORF1ab* target or only the *E* gene target was detected in a pool, an X denotes the corresponding Ct value in green or red, respectively. If both targets were detected, the average of the Ct values of both detected targets is depicted by a blue circle for that pool.

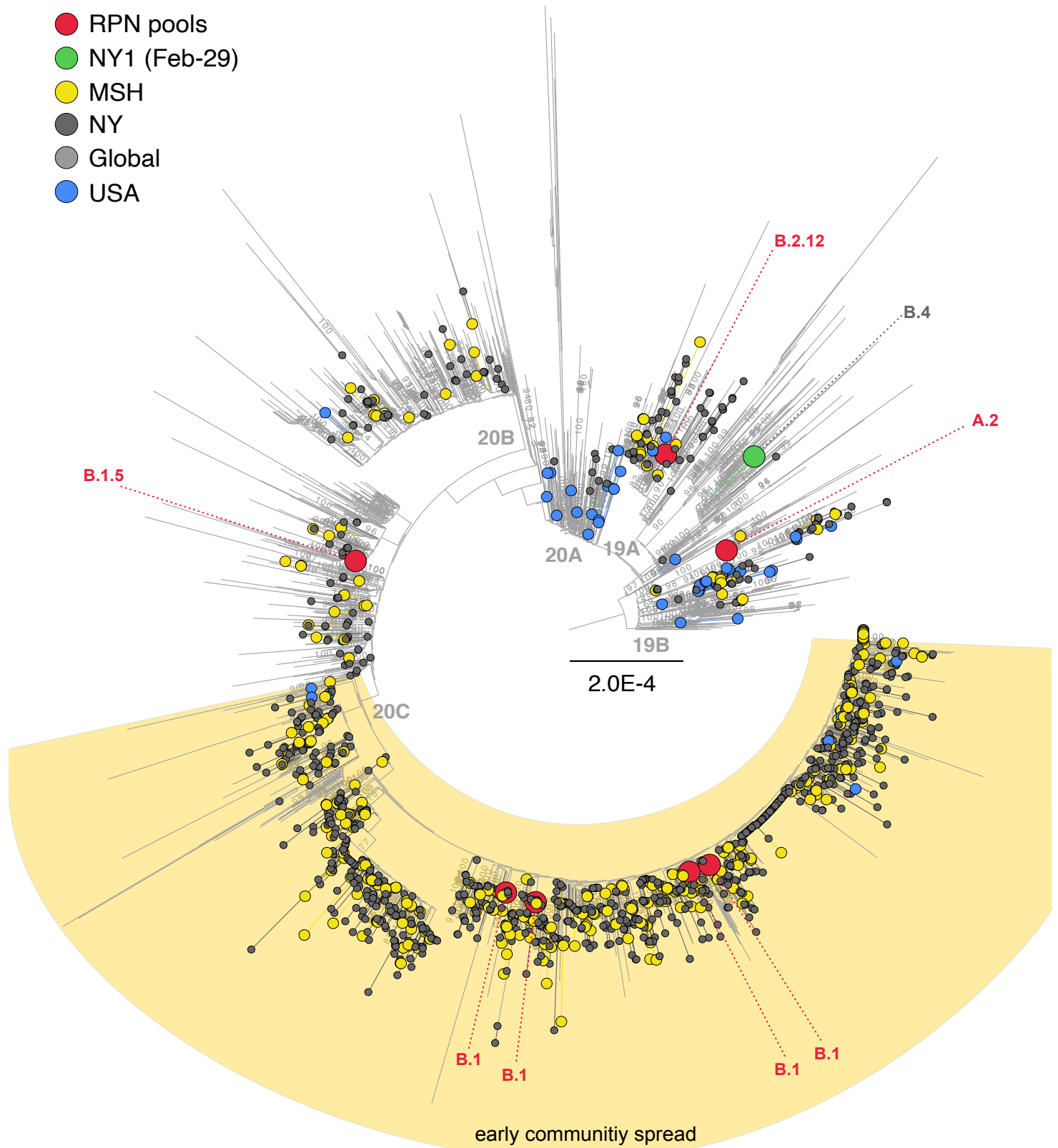
15 **Supplementary Table 1. Total and SARS-CoV-2 mapped reads per library for RNP pools with**  
 16 **incomplete genomes**  
 17

Result	RPN	Ct Values ( <i>ORF1ab</i> , <i>E</i> ) <sup>a</sup>	Library <sup>b</sup>	Total reads	SARS-CoV-2			
					Reads	%	Q30 reads	Q30 %
<i>ORF1ab</i> +	P34	36.75, ND	Nextera XT	7965566	17443	0%	2545	0.03%
			Swift	1160563	73798	6%	3355	0.29%
			Swift-low	696604	42871	6%	6430	0.92%
<i>ORF1ab</i> + <i>E</i> +	P51	34.25, 35.73	Nextera XT	738641	28298	4%	24709	3.35%
<i>ORF1ab</i> + <i>E</i> +	P58	31.66, 33.65	Nextera XT	1795845	537177	30%	505000	28.12%
			Swift	1024048	70196	7%	36608	3.57%
			Swift-low	1130048	77639	7%	62833	5.56%
<i>E</i> +	P263	ND, 36.63	Nextera XT	5187253	26720	1%	95	0.00%
			Swift	1491850	39336	3%	1905	0.13%
			Swift-low	1448665	8820	1%	121	0.01%
<i>E</i> +	P271	ND, 39.49	Nextera XT	14412083	77720	1%	448	0.00%
			Swift	2291892	118264	5%	865	0.04%
			Swift-low	1130161	61587	5%	380	0.03%
<i>E</i> +	P275	ND, 35.63	Nextera XT	5890208	44788	1%	28366	0.48%
			Swift	741823	44064	6%	333	0.04%
			Swift-low	864893	35945	4%	189	0.02%

18  
 19 a. If Ct value is not detected, “ND” is noted.  
 20 b. Nextera XT comprises data from both whole-genome and targeted custom amplicon sequencing library preparations.

21 **Supplementary Fig. 2. Maximum likelihood (ML) phylogenetic inference of seven SARS-CoV-2**  
22 **genome sequences from this surveillance study in a global background of 2,993 (same data as Fig.**  
23 **2).** Tip circles indicate the position of the respiratory pathogen-negative (RPN) pools (red) described in  
24 this report, the first reported COVID-19 case in New York City (green) from 29 February, later NYC  
25 cases from MSHS (yellow) and other institutions (dark grey), and US (blue) early isolates prior to 1  
26 March. Tips without circles correspond to the background global isolates. The yellow box delineates the  
27 position of the clade containing the majority of NYC sequences detected during the early spread. The  
28 PANGO lineage classification of the RPN pools is indicated by the dash lines, and the NextStrain clades  
29 are shown as node labels. Scale bar represents the number of nucleotide substitutions per site. The  
30 statistical support is indicated for the branches with bootstrap support values  $\geq 70\%$ .  
31

## Supplementary Fig. 2



**Supplementary Fig. 2. Maximum likelihood (ML) phylogenetic inference of seven SARS-CoV-2 genome sequences from this surveillance study in a global background of 2,993 (same data as Fig. 2).** Tip circles indicate the position of the respiratory pathogen-negative (RPN) pools (red) described in this report, the first reported COVID-19 case in New York City (green) from 29 February, later NYC cases from MSHS (yellow) and other institutions (dark grey), and US (blue) early isolates prior to 1 March. Tips without circles correspond to the background global isolates. The yellow box delineates the position of the clade containing the majority of NYC sequences detected during the early spread. The PANGO lineage classification of the RPN pools is indicated by the dash lines, and the NextStrain clades are shown as node labels. Scale bar represents the number of nucleotide substitutions per site. The statistical support is indicated for the branches with bootstrap support values  $\geq 70\%$ .

32  
33  
34

## Supplementary Table 2

### SARS-CoV-2 primer sets for whole-genome amplification

Name (1.5 Kb set)	Sequence	Pool	Tm
covid19_1.5kb_1_LEFT	ACCAACCAACTTTTCGATCTCTTGT	1	60.69
covid19_1.5kb_1_RIGHT	ATAGTGCGACCACCCTTACGAA	1	61.47
covid19_1.5kb_2_LEFT	CCACTTGCGAATTTTGTGGCAC	2	61.68
covid19_1.5kb_2_RIGHT	AGTAACCTTTGTTGGTGCACCG	2	61.58
covid19_1.5kb_3_LEFT	AGGAAGTTGTCTTGAAAACCTGGTGA	1	60.81
covid19_1.5kb_3_RIGHT	CAGCGATCTTTTGTTCAACTTGCT	1	60.79
covid19_1.5kb_4_LEFT	TCAGCACGAAGTTCTACTTGCA	2	60.22
covid19_1.5kb_4_RIGHT	AGTTGGACCAAACCTGTTGTCCA	2	60.29
covid19_1.5kb_5_LEFT	CCTATTCTGGACAATCTACACAACCTAGG	1	60.89
covid19_1.5kb_5_RIGHT	GATTATCCATTCCCTGCGCGTC	1	61.75
covid19_1.5kb_6_LEFT	TGGCTATTGATTATAAACACTACACACCC	2	61.49
covid19_1.5kb_6_RIGHT	ACTGTAGTGACAAGTCTCTCGCA	2	61.3
covid19_1.5kb_7_LEFT	GCAATTGTTTTTCAGCTATTTTGCAGT	1	60.27
covid19_1.5kb_7_RIGHT	AGTGTACTCTATAAGTTTTGATGGTGTGT	1	60.73
covid19_1.5kb_8_LEFT	ACAAGGCTATTGATGGTGGTGTCT	2	60.69
covid19_1.5kb_8_RIGHT	AGCATGTCTTCAGAGGTGCAGA	2	61.68
covid19_1.5kb_9_LEFT	GTTCTTTACCAACCACCACAAACC	1	60.68
covid19_1.5kb_9_RIGHT	GTCCACACTCTCCTAGCACCAT	1	61.48
covid19_1.5kb_10_LEFT	ACCTTCTCTTGCCACTGTAGCT	2	61.35
covid19_1.5kb_10_RIGHT	TCTGTCTAGTGCAACAGGACT	2	61.33
covid19_1.5kb_11_LEFT	TGGTACAACATTTACTTATGCATCAGC	1	60.42
covid19_1.5kb_11_RIGHT	AATACCAGCATTTTCGCATGGCA	1	61.53
covid19_1.5kb_12_LEFT	GGCATTTTGTATGAAGGTAATTGTGACA	2	61
covid19_1.5kb_12_RIGHT	AACATGTTGTGCCAACCACCAT	2	61.55
covid19_1.5kb_13_LEFT	ACGTAATGTCATCCCTACTATAACTCAA	1	60.18
covid19_1.5kb_13_RIGHT	ACAACCTGGAGCATTGCAAACA	1	61.54
covid19_1.5kb_14_LEFT	TGAGGCTATGTACACACCCGCAT	2	61.8
covid19_1.5kb_14_RIGHT	CCGAGGAACATGTCTGGACCTA	2	61.21
covid19_1.5kb_15_LEFT	TCTTTGATGAAATTTCAATGGCCACA	1	60.46
covid19_1.5kb_15_RIGHT	TGTCTGAAGCAGTGGAAAAGCA	1	60.68
covid19_1.5kb_16_LEFT	GGACTTCCTTGGAAATGTAGTGCG	2	61.48
covid19_1.5kb_16_RIGHT	CAGTGAGTGGTGCACAAATCGT	2	61.63
covid19_1.5kb_17_LEFT	AATGTAGCATTTGAGCTTTGGGC	1	60.37
covid19_1.5kb_17_RIGHT	TAACAAAGGCTGTCCACCATGC	1	61.4
covid19_1.5kb_18_LEFT	GGATCTCATTATTAGTGATATGTACGACCC	2	61.07
covid19_1.5kb_18_RIGHT	GGTCAAGTGCACAGTCTACAGC	2	61.44
covid19_1.5kb_19_LEFT	TGGTAGATTTGCCAATAGGTATTAACATCA	1	61.04
covid19_1.5kb_19_RIGHT	TGACTAGCTACACTACGTGCC	1	61.52
covid19_1.5kb_20_LEFT	AGTCCCTGTTGCTATTCATGCA	2	60.03

covid19_1.5kb_20_RIGHT	CATGACAAATGGCAGGAGCAGT	2	61.46
covid19_1.5kb_21_LEFT	AGCTGCAGAAATCAGAGCTTCT	1	60.03
covid19_1.5kb_21_RIGHT	ACATGTTCAACACCAGTGTCTGT	1	60.62
covid19_1.5kb_22_LEFT	TGCTGGCATACTAATTGTTACGACT	2	60.63
covid19_1.5kb_22_RIGHT	TCAATCTCCATTGGTTGCTCTTCA	2	60.51
covid19_1.5kb_23_LEFT	TGCTTTGCTTGTACAGTAAGTGACA	1	61.14
covid19_1.5kb_23_RIGHT	TAGCACCATAGGGAAGTCCAGC	1	61.55
covid19_1.5kb_24_LEFT	GGCCCCAAGGTTTACCCAATAA	2	60.56
covid19_1.5kb_24_RIGHT	TAGGCAGCTCTCCCTAGCATTG	2	61.61
<b>Name (2.0 Kb set)</b>	<b>Sequence</b>	<b>Pool</b>	<b>Tm</b>
covid19_2kb_1_LEFT	ACCAACCAACTTTCGATCTCTTGT	1	60.69
covid19_2kb_1_RIGHT	ACACCACCTGTAATGTAGGCCA	1	61.36
covid19_2kb_2_LEFT	GGTGCCTGGAATATTGGTGAACA	2	61.13
covid19_2kb_2_RIGHT	GCCGACAACATGAAGACAGTGT	2	61.38
covid19_2kb_3_LEFT	TGGAACTTACACCAGTTGTTTCAGAC	1	61.14
covid19_2kb_3_RIGHT	AGCATCTTGTAGAGCAGGTGGA	1	61.16
covid19_2kb_4_LEFT	AACCTCATAATTCACATGAAGGTAACA	2	60.03
covid19_2kb_4_RIGHT	AAAACACNTAAAGCAGCGGTTGA	2	61.49
covid19_2kb_5_LEFT	AAACCGTGTGTTGACTAATTATATGCCTT	1	60.13
covid19_2kb_5_RIGHT	TCACGAGTGACACCACCATCAA	1	61.86
covid19_2kb_6_LEFT	ACGAAAACAAATACGTAGTGCTGC	2	60.61
covid19_2kb_6_RIGHT	TGGGCCTCATAGCACATTGGTA	2	61.5
covid19_2kb_7_LEFT	TCTGAAGACATGCTTAACCCTAATTATGA	1	60.59
covid19_2kb_7_RIGHT	ATCACCATTAGCAACAGCCTGC	1	61.53
covid19_2kb_8_LEFT	ATTGTGGGCTCAATGTGTCCAG	2	61.14
covid19_2kb_8_RIGHT	AGCATAGACGAGGTCTGCCATT	2	61.54
covid19_2kb_9_LEFT	GCTAAATTCCTAAAACTAATTGTTGTCGC	1	60.82
covid19_2kb_9_RIGHT	TGCATTAACATTGGCCGTGACA	1	61.4
covid19_2kb_10_LEFT	CCTAAATGTGATAGAGCCATGCCT	2	60.51
covid19_2kb_10_RIGHT	TGCGAGCAGAAGGGTAGTAGAG	2	61.27
covid19_2kb_11_LEFT	TGTTGGTGATTATTTTGTGCTGACAT	1	60.24
covid19_2kb_11_RIGHT	CAAAGCACTCGTGGACAGCTAG	1	61.49
covid19_2kb_12_LEFT	CTGAGCGCACCTGTTGTCTATG	2	61.56
covid19_2kb_12_RIGHT	TGAACCTGTTTGCGCATCTGTT	2	61.59
covid19_2kb_13_LEFT	ACCCAGGAGTCAAATGGAAATTGA	1	60.45
covid19_2kb_13_RIGHT	CCTGAGGGAGATCACGCACTAA	1	61.53
covid19_2kb_14_LEFT	CAATTTTGTAAATGATCCATTTTGGGTGT	2	60.29
covid19_2kb_14_RIGHT	TGCTGCATTCAGTTGAATCACCA	2	61.31
covid19_2kb_15_LEFT	ACCCATTGGTGCAGGTATATGC	1	60.75
covid19_2kb_15_RIGHT	TGCAGTAGCGGAACAAAATCT	1	61.44
covid19_2kb_16_LEFT	CCATAGTAATGGTGACAATTATGCTTTGC	2	61.57
covid19_2kb_16_RIGHT	AGCGTTCGTGATGTAGCAACAG	2	61.49
covid19_2kb_17_LEFT	TGTGGCTCAGCTACTTCATTGC	1	61.2

covid19_2kb_17_RIGHT	TAGCACCATAGGGAAGTCCAGC	1	61.55
covid19_2kb_18_LEFT	ACTTGTACGCCTAAACGAACA	2	60.74
covid19_2kb_18_RIGHT	TAGGCAGCTCTCCCTAGCATTG	2	61.61

35

36