



**HAL**  
open science

## High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol

Nasserdine Papa Mze, Idir Kacel, Mamadou Beye, Raphael Tola, Mariéma Sarr, Leonardo Basco, Hervé Bogreau, Philippe Colson, Pierre-Edouard Fournier

### ► To cite this version:

Nasserdine Papa Mze, Idir Kacel, Mamadou Beye, Raphael Tola, Mariéma Sarr, et al.. High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. *Genes*, 2023, 14 (3), pp.681. <10.3390/genes14030681>. <hal-04115409>

**HAL Id: hal-04115409**

**<https://amu.hal.science/hal-04115409v1>**

Submitted on 6 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

## Article

# High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol

Nasserdine Papa Mze <sup>1,2,\*</sup>, Idir Kacel <sup>1,†</sup>, Mamadou Beye <sup>3,†</sup>, Raphael Tola <sup>3</sup>, Mariéma Sarr <sup>1</sup>, Leonardo Basco <sup>1</sup>, Hervé Bogreau <sup>1,4</sup>, Philippe Colson <sup>3</sup> and Pierre-Edouard Fournier <sup>1</sup>

<sup>1</sup> UMR VITROME, Aix Marseille University, IRD, AP-HM, SSA, IHU—Méditerranée Infection, 13005 Marseille, France

<sup>2</sup> Service de Biologie, Unité de Microbiologie, Hôpital Mignot, Centre Hospitalier de Versailles, 177 rue de Versailles, 78150 Le Chesnay, France

<sup>3</sup> UMR MEPHI, Aix Marseille University, IRD, AP-HM, IHU—Méditerranée Infection, 13005 Marseille, France

<sup>4</sup> Unité de Parasitologie et Entomologie, Département de Microbiologie et Maladies Infectieuses, Institut de Recherche Biomédicale des Armées, 13005 Marseille, France

\* Correspondence: npapamze@gmail.com

† These authors contributed equally to this work.

**Abstract:** The emergence of the Coronavirus Disease 2019 (COVID-19) pandemic has fostered the use of high-throughput techniques to sequence the entire severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome and track its evolution. The present study proposes a rapid and relatively less expensive sequencing protocol for 384 samples by adapting the use of an Illumina NovaSeq library to an Illumina MiSeq flow cell instrument. The SARS-CoV-2 genome sequences obtained with Illumina NovaSeq and those obtained using MiSeq instruments were compared with the objective to validate the new, modified protocol. A total of 356 (94.6%) samples yielded interpretable sequences using the modified Illumina COVIDSeq protocol, with an average coverage of 91.6%. By comparison, 357 (94.9%) samples yielded interpretable sequences with the standard COVIDSeq protocol, with an average coverage of 95.6%. Our modified COVIDSeq protocol could save 14,155 euros per run and yield results from 384 samples in 53.5 h, compared to four times 55.5 h with the standard Illumina MiSeq protocol. The modified COVIDSeq protocol thus provides high quality results comparable to those obtained with the standard COVIDSeq protocol, four times faster, while saving money.

**Keywords:** SARS-CoV-2; MiSeq; Illumina; sequencing; clade



**Citation:** Papa Mze, N.; Kacel, I.; Beye, M.; Tola, R.; Sarr, M.; Basco, L.; Bogreau, H.; Colson, P.; Fournier, P.-E. High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. *Genes* **2023**, *14*, 681. <https://doi.org/10.3390/genes14030681>

Academic Editor: Yutaka Suzuki

Received: 27 December 2022

Revised: 6 March 2023

Accepted: 6 March 2023

Published: 9 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Coronavirus Disease 2019 (COVID-19) outbreak first emerged in Wuhan, Hubei Province, People's Republic of China. It spread within a month to the rest of mainland China, then to neighboring countries, and finally to practically all countries in the world by January 2020. Since the end of the first quarter of 2020, all countries in the European Union have been affected by COVID-19 [1].

In France, the first imported cases of COVID-19 were detected on 24 January 2020 [2]. During the month of February 2020, different chains of transmission were discovered in several regions of the country. The main focus of contamination was the Oise region, situated to the north of Paris, the capital city of France. The spread of virus rapidly intensified by the end of February 2020 to reach an epidemic peak in the number of confirmed cases during the last week of March 2020. In the country, there were 38,453,595 confirmed cases of COVID-19 and 160,844 deaths, from 3 January 2020 to 10 February 2023, according to the World Health Organization (WHO) [3]. Currently in early 2023, the incidence and screening rates are decreasing in all age groups, especially among those under 10 years of age. The positivity rate is also decreasing in all age groups, except

for those aged 70 years and older. In metropolitan France, the incidence, detection, and positivity rates are decreasing in all regions. New hospitalizations due to COVID-19 are also decreasing in most regions, stable in Pays de la Loire region, and increasing in Bretagne and Centre-Val de Loire regions at low levels. The number of deaths associated with COVID-19 is also continuing to decrease, except in Bretagne [4].

The COVID-19 pandemic, which has been on-going for more than three years since early 2020, has led to the widespread use of next-generation sequencing (NGS) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes, which is essential to monitor the circulation of this virus and its evolution [5–7]. Numerous virus variants have emerged and have been identified successively, including B.1.177, B.1.160,  $\alpha$ ,  $\beta$ ,  $\gamma$ , Delta, and, since late 2021, Omicron variants [8–10]. These variants have had different degrees of impact on public health, including increased human-to-human transmissibility, disease severity, and immune escape [11,12].

According to the French National Agency of Public Health (Santé publique France) of the Ministry of Health, as of October 2022, 651,751 SARS-CoV-2 genome sequences have been characterized and deposited in the national database by French laboratories since 2021 [13]. This NGS activity has been reinforced in France through the Emergen program launched in early 2021 by the French Ministry of Health and Ministry of Research. The main objective of the Emergen program is to follow the genetic evolution of the SARS-CoV-2 virus to detect the emergence of variants and their spatio-temporal distribution in the country. SARS-CoV-2 genomics has allowed the surveillance of the emergence and spread of new variants [14]. Indeed, the predominant SARS-CoV-2 variants in France were the B.1.160 Pangolin lineage (or Marseille-4 variant) between August 2020 and January 2021, the  $\alpha$  variant since February 2021, the Delta variant since July 2021, and the Omicron variants since January 2022 [15]. At present (in early 2023), the predominant variant is BA.5.

At the University Hospital Institute-Méditerranée Infection, SARS-CoV-2 genomic surveillance began with the first diagnosis of SARS-CoV-2 infection, and variants were detected in the summer of 2020 using Oxford Nanopore MinION/GridION NGS and Illumina MiSeq then NovaSeq platforms. For the Oxford Nanopore platform, the GridION instrument is used to sequence multiple samples and track the results in real time. Using nanopore technology, a rapid sequencing protocol was developed to sequence the entire SARS-CoV2 genome in a single sample in less than two hours (first author, personal communication). However, nanopore technology is limited to sequencing a maximum of 96 samples. As for the Illumina platform, the NovaSeq instrument allows sequencing of up to 1536 samples in two flow cells of 2 lines each (i.e., 384 samples/line  $\times$  4 lines in two flow cells). By comparison, the Illumina MiSeq standard protocol is designed to sequence 96 samples simultaneously in a MiSeq flow cell.

Recently, however, it has been reported that the sequencing capacity of a modified Illumina MiSeq protocol can be increased from 96 to 384 samples in a single flow cell [6], allowing a greater throughput sequencing. In the present study, we developed a protocol that considerably increases the throughput capacity of Illumina MiSeq sequencing using a dedicated NovaSeq library for NovaSeq instrument. In a previous study conducted by Bhoyar et al. [5], it was shown that the NovaSeq library can be used to sequence 96 samples in a MiSeq flow cell. The objective of the present study was to adapt the NovaSeq library, originally designed to sequence 384 samples in the NovaSeq instrument, to sequence SARS-CoV-2 genomes from 384 samples using the MiSeq instrument in a single flow cell, thus maximizing the throughput sequencing capacity in a single run to 1536 samples (4 flow cells  $\times$  384 samples/flow cell), to compare the results obtained on the NovaSeq and MiSeq instruments.

## 2. Materials and Methods

### 2.1. Virus Samples

Nasopharyngeal swab specimens were collected from patients who presented spontaneously for routine diagnosis of SARS-CoV-2 at the University Hospital Institute-Méditerranée Infection, Marseille, southeastern France. Nasopharyngeal swabs were collected and sent to our laboratory for the diagnosis of SARS-CoV-2 infection by real-time reverse transcription-polymerase chain reaction (RT-PCR), as previously described [10]. Genome sequencing was performed on SARS-CoV-2 RNA-positive samples, as recommended by the French Ministry of Health [16]. All clinical samples were fully anonymized. The present study was reviewed and approved by the Ethics Committee of the University Hospital Institute-Méditerranée Infection under reference no. 2022-017.

### 2.2. RNA Extraction

Viral RNA extraction from nasopharyngeal swab specimens was performed based on the MVP\_2Wash\_200\_Flex protocol using an automated nucleic acid purification system (Thermo Scientific KingFisher Flex Purification System; Woodlands, Singapore), following the recommendation of the manufacturer (cat. no. A42352). Extracted RNA was stored at  $-20^{\circ}\text{C}$  until use.

### 2.3. Standard Illumina CovidSeq Protocol

Libraries were prepared following the Illumina COVIDSeq protocol (Illumina Inc, San Diego, CA, USA). There were four steps in the protocol: cDNA preparation, target amplification, library preparation and pooling, and sequencing. To prepare cDNA (first step), 8.5  $\mu\text{L}$  of extracted RNA were mixed with 8.5  $\mu\text{L}$  of elution prime fragment 3HC mix to anneal RNA. The annealed RNA was transcribed into cDNA using 1  $\mu\text{L}$  of reverse transcriptase HT. For target amplification, the synthesized cDNA was amplified (second step) by performing two multiplex PCRs with two non-overlapping primer pools (COVIDSeq Primer Pool 1 and COVIDSeq Primer Pool 2) comprising the ARTIC v3 SARS-CoV-2 specific primer set [17]. The PCR amplification conditions were as follows: initial denaturation at  $98^{\circ}\text{C}$  for 3 min, followed by 35 cycles including denaturation at  $98^{\circ}\text{C}$  for 30 s then annealing at  $65^{\circ}\text{C}$  for 5 min. Amplicons were tagged onto 4  $\mu\text{L}$  Enrichment BLT HT beads for library preparation (third step). The PCR-amplified product was processed for tagging and adapter ligation with 10  $\mu\text{L}$  of the Illumina IDT<sup>®</sup> PCR Indexes Sets 1–4 (384 indexes, for 384 samples). The labeled amplicons were amplified by PCR under the following conditions: initialization at  $72^{\circ}\text{C}$  for 3 min, initial denaturation at  $98^{\circ}\text{C}$  for 3 min, followed by 7 cycles with 20 s at  $98^{\circ}\text{C}$  for, 30 s at  $60^{\circ}\text{C}$ , 1 min at  $72^{\circ}\text{C}$ , and a final extension step at  $72^{\circ}\text{C}$  for 3 min. PCR products were batch processed in four 96-well plates. COVIDSeq HT (CPC HT) positive control and a negative control were placed in each 96-well plate. A defined volume (5  $\mu\text{L}$ ) of all 96 amplicons was pooled in a 1.5 mL microfuge tube for each plate and quantified using the Qubit dsDNA HS assay kit on a Qubit fluorometer (Invitrogen, Villebon sur Yvette, France). Each pool was diluted 1:10 with resuspension buffer, and the concentration was adjusted to 4 nM as recommended by Illumina. To perform sequencing (fourth step), 25  $\mu\text{L}$  of each pool at 4 nM were pooled in a 1.5 mL microfuge tube to constitute the final library. The library (2.25  $\mu\text{L}$ ) was diluted to a final concentration of 0.5 nM, denatured, and neutralized with 4  $\mu\text{L}$  of 0.2 N NaOH and 5  $\mu\text{L}$  of 400 mM Tris-HCl. The denatured library was mixed with 63  $\mu\text{L}$  of ExAmp mix and placed in the SP flow cell according to the NovaSeq-XP workflow (Illumina Inc). The reads length obtained was 50 bp.

### 2.4. Modified Illumina COVIDSeq Protocol

To sequence 384 samples, the final library that was obtained from four pools at 4 nM, as described above in the standard CovidSeq Illumina protocol, was used. cDNA samples (7  $\mu\text{L}$  of the library) were denatured by adding 7  $\mu\text{L}$  of 0.2 N NaOH and incubated for 5 min. A total of 986  $\mu\text{L}$  of neutralizing buffer (HT1) were added to obtain the final

volume of 1000  $\mu\text{L}$  of library. The final library (180  $\mu\text{L}$ ) was added to HT1 (420  $\mu\text{L}$ ). PhiX (6  $\mu\text{L}$ , 20 pM) was added to the latter mixture by replacing 6  $\mu\text{L}$  of the library-HT1 mixture with 6  $\mu\text{L}$  of PhiX. The library-HT1-PhiX mixture was placed in MiSeq sequencing cartridge at a final concentration of 24 pM to adjust the density to 1347  $\text{k}/\text{mm}^2$  which avoids overclustering of the flow cell (MiSeq reagent kit V2, Illumina Inc.). Sequencing was performed as recommended in the MiSeq sequencing guide (Illumina Inc.) and the reads length obtained was 250 bp.

Before sequencing 384 samples, we first sequenced 96 samples, as recommended by Bhoyar et al. [5], to obtain a cluster density of 1005  $\text{k}/\text{mm}^2$ , a passing filter of 88.3%, and Q30 quality scores of 81.6%. We sequenced 192 samples to obtain a cluster density of 1013  $\text{k}/\text{mm}^2$ , a passing filter of 91.5%, and Q30 quality scores of 81.5%. Finally, 384 amplicons were sequenced to obtain a cluster density of 1347  $\text{k}/\text{mm}^2$ , a passing filter of 81.4%, and Q30 quality scores of 76.5% (Table 1).

**Table 1.** Different protocols of library preparation to sequence amplicons using modified protocols.

Methods	Number of Amplicons Tested	Volume of Library ( $\mu\text{L}$ )	Cluster Density ( $\text{K}/\text{mm}^2$ )	Passing Filter (%)	Estimated Yield (Mb)	Q30 (G/%)
Bhoyar et al. 2021 [5]	96	5	unknown	unknown	unknown	unknown
COVIDSeq modified protocol	96 192 384	5 7 7	1005 1013 1347	88.3 92.0 81.4	8862 9274 10,244	6.9/81.6 7.2/81.5 7.5/76.5

COVIDSeq modified protocol is the method used in this study.

## 2.5. NGS Data Analysis

### 2.5.1. Standard Illumina COVIDSeq Sequencing Data

Sequence read processing and genome analysis were performed as previously described [18,19]. Briefly, base calling was carried out using the Dragen Bcl Convert pipeline v3.9.3. The bwa-mem2 tool v2.2.1 was used for mapping, which was based on Wuhan-Hu-1 isolate genome (GenBank accession no. NC\_045512.2) [20]. The resulting sequence was cleaned using SAMtools program V1.13 [21,22]. FreeBayes v1.3.5 was used for variant calling [23,24]. Consensus genomes were built with the Bcftools program v1.13 [25]. All nucleotide and deduced amino acid sequences were compared to Wuhan-Hu-1 isolate genome, and sequence modifications were detected using Nextclade tool [26–28]. Nextstrain clades and Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) were identified with the Nextclade web application [26–28] and Pangolin tool [29,30], respectively.

### 2.5.2. Modified Illumina MiSeq Sequencing Data

Sequence reading and analysis were performed as described in our previous work [4]. Trimmomatic v0.39 was used to trim raw reads and ensure quality control of sequences [31]. The alignment of the trimmed reads to the reference SARS-CoV-2 genome was performed using minimap2 (v2.17 r941) [32]. Using SAMtools (v. 1.13), the sam files of the mapping were sorted and converted into bam files [22]. A consensus sequence of the genomes in fasta format was generated using sam2consensus [33]. Sample parameters and metrics that assess the quality of the consensus genomes are shown in Supplementary Table S1.

## 2.6. Statistical Analysis

The mean values were compared using the two-sided Student *t*-test. The significance level was set at  $p < 0.05$ . All statistical tests were performed using Excel (Microsoft, Redmond, WA, USA).

### 3. Results

Of 376 samples processed for sequencing with the modified Illumina COVIDSeq protocol, 356 (94.7%) yielded interpretable genome sequences as defined by Nextclade (<https://clades.nextstrain.org/>, accessed on 2 September 2022). Of these, 280 (78.7%) yielded high-quality sequences with at least 90% coverage of the SARS-CoV-2 genome and an average sequence depth ( $\pm$  standard deviation [SD]) of  $509.6 \pm 120.94 \times$  (range, 160.0–948.1). Seventy-six samples had a coverage between 38.5% and 89.8% and a mean ( $\pm$ SD) depth of  $216 \pm 119 \times$  (range, 35.5–638.8) (Supplementary Table S2). The MiSeq genome sequences with a good quality ( $n = 329$ ) were submitted to GenBank [34] (accession numbers OP606806–OP607134).

Sequencing failed in 20 samples for which viral load was medium or low as assessed by qPCR cycle threshold values (CT) ranging from 23.0 to 33.7 cycles. For all samples from which SARS-CoV-2 genomes were sequenced with the modified COVIDSeq protocol, an average of 41.4 mutations were found. In addition, 35 different SARS-CoV-2 lineages were detected among the 356 samples that yielded interpretable genome sequences with the modified COVIDSeq protocol.

Of 376 samples sequenced using the standard Illumina COVIDSeq protocol, 357 (94.9%) yielded interpretable genome sequences as defined by Nextclade. Of these, 305 (85.4%) yielded high-quality sequences with at least 90% coverage of the SARS-CoV-2 genome and an average depth ( $\pm$ SD) of  $2988.0 \pm 563.5 \times$  (range, 795.0–4393.0). Fifty-two samples had a coverage between 45.5% and 89.5% and a mean ( $\pm$ SD) depth of  $2627.3 \pm 616.7 \times$  (range, 1187.0–4032.0) (Supplementary Table S2). Sequencing failed in 19 samples (5.1%). Sequencing also failed with these 19 samples when analyzed with MiSeq instrument. Among samples sequenced using Illumina COVIDSeq standard protocol, an average of 39.6 mutations were found. A total of 34 distinct SARS-CoV-2 lineages were detected among 357 samples that yielded interpretable sequences using the standard COVIDSeq protocol. NovaSeq genome sequences with a good quality ( $n = 344$ ) were submitted to GenBank (accession numbers OP606462–OP606805).

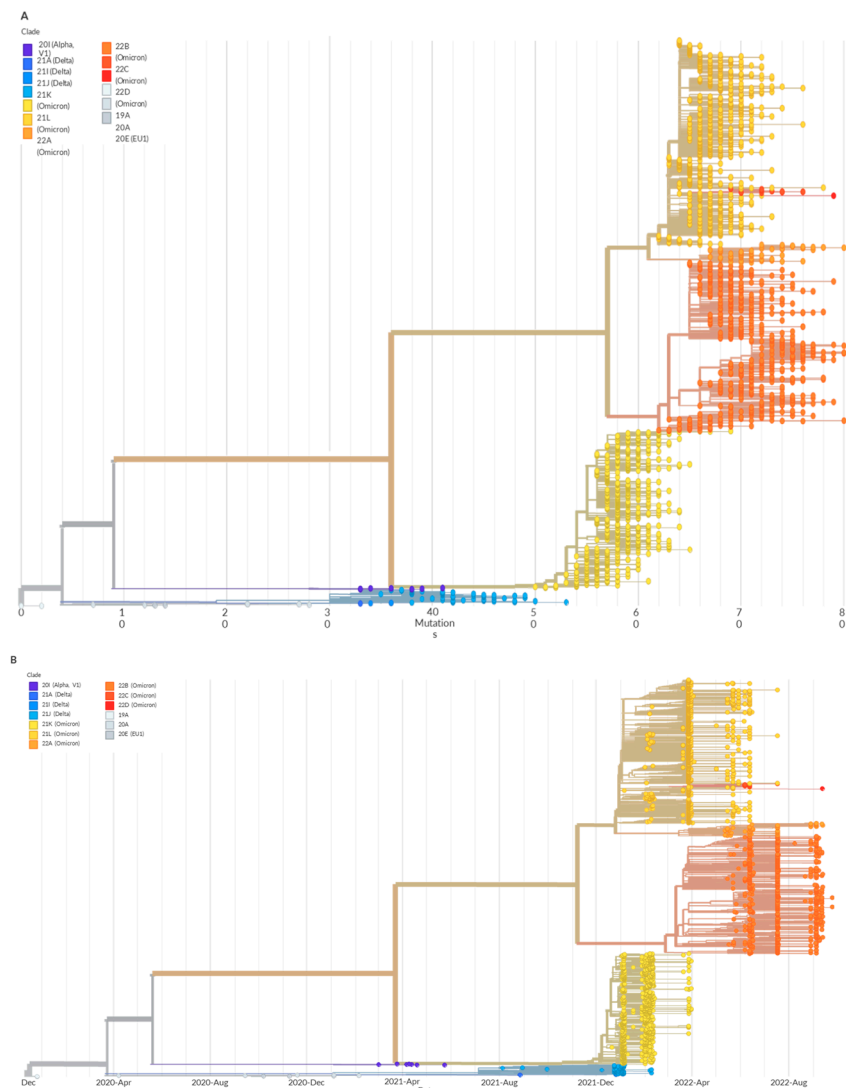
All 356 samples from which interpretable genome sequences were obtained using the modified COVIDSeq protocol were also successfully sequenced with the standard COVIDSeq protocol. Only one sample (0.3%) yielded an interpretable genome sequence (GenBank no. IHUCOVID-057737-Nova1M/2021) with the standard COVIDSeq protocol but not with the modified COVIDSeq protocol. However, the genome coverage for this sample was 49.3%, and the sequencing depth was 3045x.

For samples with SARS-CoV-2 genome sequenced successfully with both methods ( $n = 356$ ), PANGOLIN assigned the same lineage in 343 of 356 variants (96.9%) (Supplementary Table S3) and indicated a dominant occurrence of BA.1.17 ( $n = 69$ ) and BA.1.18 ( $n = 43$ ) lineages. In two samples, MiSeq consensus genome was assigned to the AY.122 variant, while the COVIDSeq consensus genome was assigned to the AY.122.6 lineage of this variant. For an additional sample, the MiSeq consensus genome was assigned to the AY.34 variant, while the COVIDSeq consensus genome was assigned to the AY.34.1 lineage. Also, there were six samples for which the MiSeq consensus genome was assigned to the BA.1 variant, while the COVIDSeq consensus genome was assigned to the BA.1.1 lineage. Sequencing depth was higher for sequences obtained using the COVIDSeq protocol (mean  $\pm$  SD,  $2935.7 \pm 584.7$ ) compared to those obtained with the modified COVIDSeq protocol ( $446.8 \pm 170.3$ ), and the difference was significant ( $p = 0.0001$ ). Coverage was similar between the two methods (mean  $\pm$  SD,  $94.6 \pm 9.7\%$  versus  $91.6 \pm 11.9\%$  for the COVIDSeq protocol and the modified protocol, respectively;  $p = 0.84$ ). There were more mutations detected (41.3 versus 37.4 on average) with the modified COVIDSeq protocol, but the difference was not significant ( $p = 0.74$ ). Otherwise, both protocols assigned the variants to the same clades based on the nomenclature defined by Nextstrain, suggesting that a majority (i.e., 279 of 356 [78.4%] genomes) of the samples belonged to clade 21K while 65 (18.2%) belonged to clade 21J. Four variants belonged to clade 21I, two to clade 20A, one to clade 21L, and five to clade 20I (Table 2).

**Table 2.** Comparison of SARS-CoV-2 clades monitored with two sequencing protocols.

Technique	20A	20I	21I	21J	21K	21L	Total
MiSeq	2	5	4	65	279	1	356
NovaSeq	2	5	4	66	279	1	357

Therefore, our results strongly suggest that high quality genomic sequences can be obtained by multiplexing 384 NovaSeq libraries in a single MiSeq flow cell. This methodological approach allowed us to genotype 3384 samples to detect mutation patterns, identify SARS-CoV-2 lineages, and follow their chronological evolution (Figure 1A,B).



**Figure 1.** Phylogenetic distribution of SARS-CoV-2 genomes. (A) Phylogenetic trees generated by Nextstrain [29]. Of 3384 genomes, 2519 are highlighted. Clade 22 B is the dominant clade. This phylogenetic analysis shows different variants present in all sequences and their mutational profiles. (B) Phylogenetic trees showing the evolution of SARS-CoV-2 variant strains.

Regarding the financial cost of the two protocols, it was shown that the estimated cost to sequence 384 samples was approximately 6794 euros using the modified COVIDSeq protocol, as compared to 8025 euros using the standard COVIDSeq protocol (Table 3). Therefore, the standard COVIDSeq protocol was more expensive. Regarding the time of run (without including sequence analysis time), it was 22 h with the standard COVIDSeq protocol, as compared to 52 h with the modified COVIDSeq protocol.

**Table 3.** Comparison of financial costs and required estimated time for two sequencing protocols.

Technique	Reagent	Manufacturer	Total Duration (h)	Cost (euros)	Total Cost (euros)	
					NovaSeq	MiSeq
protocol modified for COVIDSeq and MiSeq (for preparation of library)	Illumina COVIDSeq™ Test (384 Samples)	Illumina	8	4212	8025	6794
	IDT for Illumina PCR Indexes Set 1–4	Illumina		1267		
	Illumina COVIDSeq v4 Primer Pools, 384 Samples RUO	Illumina		235		
	Qubit dsDNA HS Assay Kit	Thermo Fisher		60		
NovaSeq	NovaSeq 6000 SP Reagent Kit (100 cycles) V1.5	Illumina	14	1971		
	NovaSeq XP 2-Lane Kit V1.5	Illumina		280		
MiSeq	MiSeq Reagent Kit V2 (500 cycles)	Illumina	44	1020		

#### 4. Discussion

Researchers worldwide are being encouraged by international organizations and governments to collect and sequence SARS-CoV-2 genomes and share sequence data through platforms like GenBank or Global Initiative on Sharing Avian Influence Data (GISAID) where currently (as of February 2023) more than 14,919,562 genome sequences from many countries are available [35]. Databases with sequences from different periods of time and laboratories around the world would facilitate the monitoring of SARS-CoV-2 and the understanding of its past, current, and potential evolutionary changes. Therefore, the availability of relatively simple, rapid, high throughput and the cheapest possible sequencing methods is of great interest.

Recently, a new protocol for sequencing SARS-CoV-2 genomes using a COVIDSeq library was described [5]. Such a library is usually used to perform NGS on a NovaSeq instrument that is very expensive (about one million euros). The optimized protocol used a COVIDSeq library on a MiSeq instrument that is far less expensive (approximately four times) than the NovaSeq instrument. However, in the study conducted by Bhoyar et al. [5], only NGS was performed for 96 samples in a flow cell, which means that it would take a much longer time and would need four flow cells to sequence 384 amplicons. In the present study, we developed and validated a simple, fast, and cheaper method to sequence 384 SARS-CoV-2 genomes in a single run, in a single flow cell, on a MiSeq instrument. Minor modifications to the protocol developed by Bhoyar et al. [5] were necessary to achieve the NGS of 384 samples in a single flow cell.

The modified Illumina COVIDSeq protocol provided high quality sequences in our experience despite the increased number of samples to be processed simultaneously. A greater sequencing depth was obtained with the standard COVIDSeq protocol than with our modified COVIDSeq protocol, which might be the result of the fact that the performance of the NovaSeq instrument produces a large amount of high-quality sequences. Nevertheless, the final genome sequences of various SARS-CoV-2 lineages obtained using both protocols were identical for all. Interestingly, at lower, but still substantial sequencing depth, more mutations were identified using the modified COVIDSeq protocol.

In the present study, we have shown that it is possible to sequence amplicons obtained from 384 SARS-CoV-2-positive samples in a single flow cell by slightly modifying the standard COVIDSeq protocol. Our modified protocol allowed to sequence SARS-CoV-2 genomes from 3384 samples, contributing to the monitoring of the circulation and evolution of SARS-CoV-2 variants in southern France. As a matter of fact, in practice, this modified COVIDSeq protocol was used as a backup method when our NovaSeq instrument was not operational. Hence, this strategy contributed efficiently to local and national SARS-CoV-2

genomic surveillance. Furthermore, if we compare the cost to perform the modified COVID-Seq protocol for NGS on the MiSeq instrument to the one that our team previously also reported for NGS on the MiSeq instrument [4], the total cost to sequence 384 samples was estimated to be 6794 euros with the modified COVIDSeq protocol compared to 15,380 euros with the modified MiSeq protocol. Therefore, based on our cost–benefit estimation, the modified COVIDSeq protocol saves 9603 euros per 384 samples. This difference is largely due to the fact that Illumina COVIDSeq reagent is less expensive than the MiSeq reagent, especially when purchased to sequence large numbers of samples (e.g., 33,700 euros for 3072 samples). Moreover, it takes 52 h (without the time required for sequence analysis) to sequence 384 samples regardless of the protocol (modified MiSeq or modified COVIDSeq protocols). If we compare the standard MiSeq protocol, the cost of which has been estimated to be 21,000 euros for 55.5 h [4], to the modified CovidSeq protocol, the latter saves approximately 15,223 euros and 2 h.

## 5. Conclusions

The standard COVIDSeq Illumina protocol can be modified to sequence 384 amplicons simultaneously in a single flow cell without compromising the high-quality sequences. This modified protocol is less expensive, more rapid, and compatible with the Illumina MiSeq sequencer. This approach may be a promising SARS-CoV-2 genome NGS strategy, notably in a situation when a NovaSeq instrument is not available or operational. This modified protocol could be used in countries that only have the MiSeq sequencer, especially in low-income countries, which could allow them to conduct surveillance for SARS-CoV-2 variants at a lower cost.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14030681/s1>, Table S1: Data generated from the modified Illumina COVIDSeq protocol; Table S2: Data generated from the Illumina COVIDSeq procedure; Table S3: Data generated for comparison between data from the modified Illumina COVIDSeq protocol and from the Illumina COVIDSeq procedure.

**Author Contributions:** N.P.M. performed experiments and wrote the draft of the manuscript. P.C. and P.-E.F. conceived and designed the study. M.B. assisted with analysis, edited, and reviewed the manuscript. I.K., M.S., R.T., H.B. and L.B. contributed to data analysis and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the French Government under the “Investments for the Future” programme managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03 and was also supported by Région Provence Alpes Côte d’Azur and European funding PRIMMI (Plateformes de Recherche et d’Innovation Mutualisées Méditerranée Infection), FEDER PA 0000320 PRIMMI.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the Mediterranean Infection institute under reference (No. 2022-017).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The high-quality genomic sequences were submitted to the GenBank (accession numbers OP606806-OP607134 for the modified Illumina COVIDSeq protocol and OP606462-OP606805 for the Illumina COVIDSeq protocol).

**Acknowledgments:** We thank the team in charge of SARS-CoV-2 RNA extraction and genomics at IHU–Méditerranée Infection, especially Laurence Thomas, Ludivine Brechard, Aylin Karadeniz, Sarah Durand, Lisa Biscioni, and Claudia Andrieu.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

COVID-19: coronavirus disease 2019; C<sub>T</sub>, cycle threshold; GISAID, Global Initiative on Sharing Avian Influenza Data; Mbp, megabase; NGS, next generation sequencing; PAN-GOLIN, Phylogenetic Assignment of Named Global Outbreak Lineages; pb, base pair; RSB, Resuspension Buffer; RT-PCR, reverse transcription-polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SD, standard deviation; WHO, World Health Organization.

## References

1. World Health Organization. Listings of WHO's Response to COVID-19. Available online: <https://www.who.int/news/item/29-06-2020-covid-timeline> (accessed on 2 March 2023).
2. Stoecklin, S.B.; Rolland, P.; Silue, Y.; Mailles, A.; Campese, C.; Simondon, A.; Mechain, M.; Meurice, L.; Nguyen, M.; Bassi, C.; et al. First cases of coronavirus disease 2019 (COVID-19) in France: Surveillance, investigations and control measures, January 2020. *Euro Surveill.* **2020**, *25*, 2000094.
3. World Health Organization. France: WHO Coronavirus Disease (COVID-19) Dashboard with Vaccination Data. Available online: <https://covid19.who.int/region/euro/country/fr> (accessed on 13 February 2023).
4. Santé Publique France. InfoCovidFrance: Chiffres clés et Evolution de la COVID-19 en France et dans le Monde. Available online: <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolution-de-la-covid-19-en-france-et-dans-le-monde> (accessed on 13 February 2023).
5. Bhojar, R.C.; Senthivel, V.; Jolly, B.; Imran, M.; Jain, A.; Divakar, M.K.; Scaria, V.; Sivasubbu, S. An Optimized, Amplicon-Based Approach for Sequencing of SARS-CoV-2 from Patient Samples using COVIDSeq Assay on ILLUMINA MiSeq Sequencing Platforms. *STAR Protoc.* **2021**, *2*, 100755. [[CrossRef](#)] [[PubMed](#)]
6. Papa Mze, N.; Beye, M.; Kacel, I.; Tola, R.; Basco, L.; Bogreau, H.; Colson, P.; Fournier, P.E. Simultaneous SARS-CoV-2 genome sequencing of 384 samples on an Illumina MiSeq instrument through protocol optimization. *Genes* **2022**, *13*, 1648. [[CrossRef](#)] [[PubMed](#)]
7. Fuchs, M.; Radulescu, C.; Tang, M.; Mahesh, A.; Lavin, D.; Umbreen, S.; McKenna, J.; Smyth, M.; McColgan, E.; Molnar, Z.; et al. Mini-XT, a miniaturized tagmentation-based protocol for efficient sequencing of SARS-CoV-2. *J. Transl. Med.* **2022**, *20*, 105. [[CrossRef](#)] [[PubMed](#)]
8. Fournier, P.E.; Colson, P.; Levasseur, A.; Devaux, C.A.; Gautret, P.; Bedotto, M.; Delerce, J.; Brechard, L.; Pinaul, L.; Lagier, J.C.; et al. Emergence and outcomes of the SARS-CoV-2 'Marseille-4' variant. *Int. J. Infect. Dis.* **2021**, *106*, 228–236. [[CrossRef](#)] [[PubMed](#)]
9. Hodcroft, E.B.; Zuber, M.; Nadeau, S.; Vaughan, T.G.; Crawford, K.H.D.; Althaus, C.L.; Reichmuth, M.L.; Bowen, J.E.; Walls, A.C.; Corti, D.; et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **2021**, *595*, 707–712. [[CrossRef](#)] [[PubMed](#)]
10. World Health Organization. Tracking SARS-CoV-2 Variants. Available online: <https://www.who.int/activities/tracking-SARS-CoV-2-variants> (accessed on 12 October 2022).
11. DeGrace, M.M.; Ghedin, E.; Frieman, M.B.; Krammer, F.; Grifoni, A.; Alisoltani, A.; Alter, G.; Amara, R.R.; Baric, R.S.; Barouch, D.H.; et al. Defining the risk of SARS-CoV-2 variants on immune protection. *Nature* **2022**, *605*, 640–652. [[CrossRef](#)] [[PubMed](#)]
12. Davies, N.G.; Abbott, S.; Barnard, R.C.; Jarvis, C.I.; Kucharski, A.J.; Munday, J.D.; Pearson, C.A.B.; Russell, T.W.; Tully, D.C.; Washburne, A.D.; et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **2021**, *372*, eabg3055. [[CrossRef](#)] [[PubMed](#)]
13. Santé Publique France. Consortium EMERGEN—Activité Hebdomadaire de Séquençage. Available online: [www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-circulation-des-variants-du-sars-cov-2#block-338801](http://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-circulation-des-variants-du-sars-cov-2#block-338801) (accessed on 11 September 2022).
14. Hodcroft, E. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. 2021. Available online: <https://covariants.org/> (accessed on 11 September 2022).
15. Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.; Bleicker, T.; Brunink, S.; Schneider, J.; Schmidt, M.L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *EuroSurveillance* **2020**, *25*, 2000045. [[CrossRef](#)] [[PubMed](#)]
16. Santé Publique France. Consortium EMERGEN. Available online: <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/consortium-emergen> (accessed on 27 September 2022).
17. ARTIC Network. Real-Time Molecular Epidemiology for Outbreak Response. Available online: <https://artic.network> (accessed on 11 September 2022).
18. Colson, P.; Fournier, P.E.; Chaudet, H.; Delerce, J.; Giraud-Gatineau, A.; Houhamdi, L.; Andrieu, C.; Brechard, L.; Bedotto, M.; Prudent, E. Analysis of SARS-CoV-2 variants from 24,181 patients exemplifies the role of globalization and zoonosis in pandemics. *Front. Microbiol.* **2022**, *12*, 786233. [[CrossRef](#)] [[PubMed](#)]

19. Colson, P.; Fournier, P.E.; Delerce, J.; Million, M.; Bedotto, M.; Houhamdi, L.; Yahi, N.; Bayette, J.; Levasseur, A.; Fantini, J.; et al. Culture and identification of a “Deltamicron” SARS-CoV-2 in a three cases cluster in southern France. *J. Med. Virol.* **2022**, *94*, 3739–3749. [[CrossRef](#)] [[PubMed](#)]
20. GitHub. v2.2.1. Available online: <https://github.com/bwa-mem2/bwa-mem2> (accessed on 12 September 2022).
21. SAMtools. Available online: <https://www.htslib.org> (accessed on 12 September 2022).
22. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [[CrossRef](#)] [[PubMed](#)]
23. GitHub. Available online: <https://github.com/freebayes/freebayes> (accessed on 12 September 2022).
24. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907.
25. BCFtools. Available online: <https://samtools.github.io/bcftools/bcftools.html> (accessed on 12 September 2022).
26. Nextclade Web 2.12.0, Nextclade CLI 2.12.0. Available online: <https://clades.nextstrain.org> (accessed on 12 September 2022).
27. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *34*, 4121–4123. [[CrossRef](#)] [[PubMed](#)]
28. Aksamentov, I.; Roemer, C.; Hodcroft, E.B.; Neher, R.A. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **2021**, *6*, 3773. [[CrossRef](#)]
29. Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN). Available online: <https://cov-lineages.org/pangolin.html> (accessed on 12 September 2022).
30. Rambaut, A.; Holmes, E.C.; O’Toole, A.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403–1407. [[CrossRef](#)]
31. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
32. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **2021**, *37*, 4572–4574. [[CrossRef](#)] [[PubMed](#)]
33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
34. Sayers, E.W.; Cavanaugh, M.; Clark, K.; Pruitt, K.D.; Schoch, C.L.; Sherry, S.T.; Karsch-Mizrachi, I. GenBank. *Nucleic Acids Res.* **2022**, *50*, D161–D164. [[CrossRef](#)] [[PubMed](#)]
35. Global Initiative on Sharing Avian Influenza Data (GISAID). Available online: <https://gisaid.org/> (accessed on 12 September 2022).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.